


```

360 .....LysThrSerThrIleThrGlyGlyLysSerLeuAsn 372
1081 GAAACTGATTAAGAACAGTATTAGCGGCGAGGGGTGTTATTCAGTACCG 1130
373 .....ValAspLeuAlaAspGly.....Lys 379
1131 TCCAAGTTAAACAGAGGTGAAGAACTTTCTTTATTCGATTACGCAACG 1180
379 sasprysProAsnHisGlyLysSerValThrPhe.....GluGlySerG 394
1181 GCAAACTCATCTTATCAACACATCAACACGAGCGCGCGGTGTGAT 1230
394 lYThrLeuThrLeuAsnAsnHisIleAspGlnGlyAlaGlyLysLeuPhe 410
1231 TTGAGGCGATTTTACGTC...TCGCGCTGAAACAGCAAGAACGTGCA 1277
411 PheGluGlyAspTyrGluValLysGlyThrSerAspAsnThrThrTrrPly 427
1278 AGCGCGGCGGCTTCATATCATGAGACAGTACCGTTACTTGAAGATTA 1327
427 sGlyAlaGlyValSerValAlaGluGlyLysThrValThrTrrPlyValH 444
1328 AGCGCGTGCAACAGACCGCGCTGTCCAAATCGGCAAGGACGCTGCAC 1377
444 lAsnProGlnTyrAspArgLeuAlaLysIleGlyLysGlyThrLeuIle 460
1378 GTTCAAGCCAAAGGGGAAACCAAGCGTGCATCAGCGTGGGAGCGGTAC 1427
461 ValGluGlyThrGlyAspAsnLysGlySerLeuLysValGlyAspGlyTh 477
1428 AGCTATTTTGGATCAGACGACAGAGATTAAGGCAAAAGACCTTTA 1477
477 rValIleLeuLysGlnGlnThrAsnGlySerGly...GlnHisAlaPheA 493
1478 GTGAATCGGCTTGTCAGCGGCGAGCGGTACGTCGAATGATCCGAT 1527
493 lAspValGlyIleValSerGlyArgSerThrLeuValLeuAsnAspAsp 509
1528 AATCGATTACCCCGACAACTATTTGCGCTTTCGCGGCGGACGTTT 1577
510 LysGlnValAspProAsnSerIleTyrPheGlyPheArgGlyGlyArg 526
1578 GGATTTAAACGGGCATTCGCTTCGTCACCGTATTCAAATACCGATG 1627
526 uAspLeuAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspG 543
1628 AAGGGCGCATGTGNCATCATATGCCACACACATCCACGCTTACC 1677
543 lGlyAlaArgLeuValAsnHisSerThrSerLysHisSerThrValThr 559
1678 ATTACAGGGAATGAAGATTTACACACGAGTGGTAGAATATC..... 1722
560 lIleThrGlyAspAsnLeuIleThrAspProAsnAsnValSerIleTyr 576
1722 ..... 1722
576 rValLysProLeuGluAspAspAsnProTyrAlaIleArgGlnIleLysT 593
1722 ..... 1722
593 yrcGlyTyrGlnLeuTyrPheAsnGlnGluAsnArgThrTyrTyrAlaLeu 609
1722 ..... 1722
610 LysLysAspAlaSerIleArgSerGluPheProGlnAsnArgGlyGluSe 626
1722 ..... 1722
626 rAsnAsnSerTrrPleuTyrMetGlyThrGluLysAlaAspAlaGlnLysA 643
1723 .....AATAGACTTAATTACAGCAAGAATTCCTACACAGCGTTGG 1764
643 snAlaMetAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyr 659

```

```

1765 TTTCGGGAGAAAGATAGACCAAAAGACGGCGGCTCAACCTTGTTA 1814
660 PheGlyGluGluGlu...GlyLysAsnAsnGlyAsnLeuAsnValThrPh 675
1815 CCAGCCCGCCGACGAAGACCGCACCCNGCTGCTTCCGCGGCAACAAT 1864
675 eLysGlyLysSerGluGlnAsnArgPheLeuLeuIrrGlyGlyThrAsnL 692
1865 TAAACGGCACATCACGCCAAACAAACGGCAAACTGTTTTCACGCGCAG 1914
692 euAsnGlyAspLeuAsnValGlnGlnGlyThrLeuPheLeuSerGlyArg 708
1915 CCGACACCGCAGCGCTACAAATCATTTAGAAAGCGGTGTCAAAAATGA 1964
709 ProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAs 725
1965 AGGT.....ATCCACACAGAGAAATCGTGTGGACACAGCATGATCN 2008
725 pSerHisPheSerGluAsnAsnGlnValValAlaValAspAspTrrPleA 742
2009 ACCGCACTTTAAAGCGGAAATTTCCATATTCAGGCGGCGGCGGTG 2058
742 snArgAsnPheLysAlaThrAsnIleAsnValThrAsnAlaThrLeu 758
2059 ATTTC...CGCAATGTTGCCAAAGTGGAAGCGGATTCGATTTGAGCAA 2105
759 TyrSerGlyArgAsnValGluSerIleThrSerAsnIleThrAlaSerAs 775
2106 TCACGCCCAAGCAGTTTGGTGTGCGACCGCATCAAGCCATACATCT 2155
775 nAsnAlaLysValHisIleGlyTyr....LysAlaGlyAspThrValC 790
2156 GTACAGCTGGAGTGCAGNGGTCTGACAAATGTGTCGAANAANAAT 2205
790 yValAlaArgSerAspTrrThrGlyTyrValThrCysTrrThrAspLysLeu 806
2206 ACCGACGATAAGTATGCTTCATTCATGACTAAGACACNTNMGCGCAN 2255
807 Ser...AspLysAlaLeuAsnSerPheAsnProThr..... 817
2256 TGTNAGCTNCCNATNACNTNNTNNAANCCTCNCNGGCGCTGCNNCAC 2305
817 ..... 817
2306 TNAANGCAATCTTAGTCAAAATGGCGATACAGTTATACAGTACGCCAC 2355
818 .....AsnLeu..... 819
2356 AACGCCACCCAAACGGCAACCTTAGCGCTGTGGCAATGCCCAAGCAAC 2405
820 .....ArgGlyAsnValAsnLeuThrCylSerIleAsnPheVa 832
2406 ATTAAATCAACCCACATTAACGGCAACNATCGGATTCGGGCAATGCTT 2455
832 lLeuGlyLysAlaAsnLeuPheGlyThrIleGlnSerArgGlyAsnSerG 849
2456 CATTAACTTAGCAACACAGCGCGCACAAAGGACAGTGTGAGCGCTTCC 2505
849 lValArgLeuThrGluAsn..... 855
2506 GACAACGCTAAGCAAAAGTAAAGCATTCGCGACTCAAGGCAATGTCTC 2555
855 ..... 855
2556 CCTAGCCGATAGGACAGATTCATTGTAACACGCCGCTTTACCGGAC 2605
855 ..... 855
2606 AACTCAGCGCAGCAGAGANACAGCATTTACACTTAAAGACAGGAGTGG 2655
856 .....SerHisTrr 858

```



```
492 TAAATTTGTACAGATGCAAGAACTGTGCA...ATGACGAGTCAGCATGA 538
|||||
171 pypSpheValThrGluValAlaProIleGluIaSerThrAlaSerSera 188
|||||
539 GGGGAAATCCATTCGCCATAAAGAAAATATCCGCGAGTGTCCGCAATC 588
|||||
188 spIaGlyThrTyraSpGlnAsnLysTyProAlaPheValArgLeu 204
|||||
589 GCGTCAGACACCACTATGCGCTATGATGATGACAAAACAGCGCATTT 638
|||||
205 GlySerGlySerGlnPheIleTy...LysGlyGlyAsp... 216
|||||
639 ATCTACTCCGCGCGCATGTTATTTGGCGGCAATACATATGACGGTT 688
|||||
217 .AsnTySer...LeuIleLeuAsnAsnHisGluValGly... 228
|||||
689 GGGGAAATTAAT...GGCGTANTT 708
|||||
229 ..GlyAsnAsnLeuLysLeuValGlyAspAlaTyThrTyGlyIleAla 244
|||||
709 AGTTTGAGCGCGCATGTCCGCGCATGCAACGAC...TATGG 746
|||||
245 GlyThrProTyThrValAsnHisGluAsnAsnGlyLeuIleGlyPheG 261
|||||
747 C.....C 748
|||||
261 yAsnSerLysGluGlnHisSerAspProLysGlyIleLeuSerGlnAsp 278
|||||
749 CTATGCGCATTCGAGGTGCGGCGGACAGCGGTTCCGCAATGTTTNT 798
|||||
278 roLeuThrAsnTyThrAlaValLeuGlyAspSerGlySerProLeuPheVal 294
|||||
799 TATGACAAAACAACAATTAATGCTGCTCAACGAGTTTACAAACCGG 848
|||||
295 TyraSpArgGlnLysGlyLysThrLeuPheLeuGly.....Se 307
|||||
849 CTACCTTATTCGCGAGGAAACGGTTTCCAGCTGATACGCAAGATT 898
|||||
307 tTyraSpPheThrAlaGlyTyraSpn.....LysLysSer 319
|||||
899 GGTTCAGCATGACATTTACAGAGCGATACACATACCGTCTNTTTGAA 948
|||||
319 rGlnGlnIuThrAsnIleTyLysSer..... 327
|||||
949 CGCGCGAGTACGACATTTTCTTTTACATCC.....ACACAGAA 989
|||||
328 .....GlnPheThrLysAspValLeuAsnLysAs 337
|||||
990 CGGTACGGGTACGGTACAGAAACCAAGAAAGTNTCC...AATCCAA 1036
|||||
337 pSerIaGlySerLeuIleGlySerLysThrAspTySerThrSerSera 354
|||||
1037 ACCTTAAGTACAGACAGTCGACTGTTTACAGATCTTTGAATAAACT 1086
|||||
354 snGlyLysThrSerThrIleThrGlyGlyGluLysSerLeuAsn..... 368
|||||
1087 GATAAAGAACAGTTTACCGCGGAGGGGTGTTAATCAGTACCGCTCAAG 1136
|||||
369 .....ValAspLeuAlaAspGly.....LysAspLys 377
|||||
1137 GTTAAACAACGGTGAACCTTTCTTTATCATGATACGCAACGGCAAC 1186
|||||
377 pProAsnHisGlyLysSerValThrPhe.....GluGlySerGlyThr 392
|||||
1187 TCATTTATCAACAACATCAACCAAGCGCGGGGCTTTGATTTTGA 1236
|||||
392 euThrLeuAsnAsnAsnIleAspGlnGlyAlaGlyGlyLeuPhePheGlu 408
|||||
1237 GGTGATTTTACGGTC...TCGCGTGAACAAACGAACGTGGCAAGCGC 1283
|||||
409 GlyAspTyThrGluValLysGlyThrSerAspAsnThrThrTyLysGlyAl 425
|||||
1284 GGGCGTTCATATACAGTGAAGACATACCGTTACTTGGAAGTAAACGGCG 1333
|||||
425 aglyValSerValAlaGlnGlyLysThrValThrTyLysValHisAsn 442
|||||
1334 TCGCAACACCGCGCTGCCAAAATCGGCAAGGACGCTGACGTCATCA 1383
|||||
442 roGlnTyraSpArgLeuAlaLysIleGlyLysGlyThrLeuIleValGlu 458
|||||
1384 GCCAAAGGGGAAAAACCAAGGCTGATCAGCGTGGCGGACGTCATCAT 1433
|||||
459 GlyThrGlyAspAsnLysGlySerLeuLysValGlyAspGlyThrVal 475
|||||
1434 TTTGGATCAGCAGGACAGCATTAAGGCAAAAAACAAGCTTATAGTGA 1483
|||||
475 eLeuLysGlnGlnThrAsnGlySerGly...GlnHisAlaPheAlaSer 491
|||||
1484 TCGCGTTCGNTCAGCGGCGGATACGCTGCACTGCAATGCGCATTA 1533
|||||
491 alGlyIleValSerGlyArgSerThrLeuValLeuAsnAspAspLysGln 507
|||||
1534 TTCAACCCGCAACAACCTATTTCCGCTTTCGCGGCGGACGTTGGATT 1583
|||||
508 ValAspProAsnSerIleTyThrPheGlyPheArgGlyGlyArgLeuAsp 524
|||||
1584 AAACGGCATTCGCTTTCGTTCCACCGTATTCAAAATACGATGAAGGG 1633
|||||
524 uAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspAspGly 541
|||||
1634 CGATGATTCGTCATATGTCACACACATCCACCGTTTACCATTA 1683
|||||
541 IaArgLeuValAsnHisAsnMetThrAsnAlaSerAsnIleThr 557
|||||
1684 GCGAATGAAGATTTACACACCGAGT..... 1710
|||||
558 GlyLysSerLeuIleThrAspProAsnThrIleThrProTyraAsnLea 574
|||||
1710 ..... 1710
|||||
574 pAlaProAspGluAspAsnProTyraAlaPheArgArgIleLysAspGly 591
|||||
1710 ..... 1710
|||||
591 LysGlnLeuTyThrLeuAsnLeuGluAsnTyThrTyThrAlaLeuArgLys 607
|||||
1710 ..... 1710
|||||
608 GlyAlaSerThrArgSerGluLeuProLysAsnSerGlyGluSerAsnG 624
|||||
1711 .....GTAAGAAT.....A 1720
|||||
624 uAsnTyThrLeuTyThrGlyLysThrSerAspGluAlaLysArgAsnVal 641
|||||
1721 TCATAGACTTAATTTACAGCAAGAAATTTGCTTACACAGGTTGTTGGC 1770
|||||
641 eAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyThrPheGly 657
|||||
1771 GAGAAAGATACGACCAACCAAGGCGGCTCAACCTGTTTACAGGCC 1820
|||||
658 GlnGlnGln...GlyLysAsnAsnGlyAsnLeuAsnValThrPheLysG 673
|||||
1821 CGCGCGAGACCGCCACCCGCTGCTTCCGCGGCAACAAATTTAAGC 1870
|||||
673 LysSerGlnGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAsnG 690
|||||
1871 GCAACATCGCAACAAACAGGCAAACTGTTTTCAGCGGCGACAGCGACA 1920
|||||
690 LysAspLeuThrValGluLysGlyThrLeuPheLeuSerGlyArgProThr 706
|||||
1921 CGGACCGCTTACATCATTTAGAGAGCGGTGCTCAAAAATGGAAGGT.. 1968
|||||
707 ProHisIleArgAspIleAlaGlyIleSerSerThrLysLysAspProH 723
|||||
1969 ....ATCCCAAGGAGAAATCGTGGCGACACGACTGGATGNACCGCA 2014
|||||
```

```
723 sPheLaGluAsnAsnGluValValGluAspSerPheLeuAsnArg 740
2015 CGTTTAACCGGAAATTCATTCAGAGCGCGCGGTGATTCC 2064
:::|||||:::|||||:::|||||:::|||||
740 snPheLysAlaThrMetAsnValThrGlyAsnAlaSerLeuTyr 756
2065 ...CGCAATGTTGCCAAAGTGGAGCGCATGTCATTGAGCAATCAGC 2111
|||||:::|||||:::|||||:::|||||
757 GlyAlaGlnValAlaAsnIleThrSerAsnIleThrAlaSerAsn 773
2112 CCAAGCAGTTTTGTGTGCGACCGCATCAAGCCATCAATCTGTAC 2161
|||||:::|||||:::|||||:::|||||
773 agInValHisIleGlyTyr...LysThrGlyAspThrValGlyVal 788
2162 GTTGGCAGTGCAGCGCTGCAATGTGTGTGCAAAANCAATTAhCC 2211
|||||:::|||||:::|||||:::|||||
788 rGSerAspTyrThrGlyTyrValThrCysThrThrAspLysLeu 803
2212 GATTAAGTATGCTTCATTCAGTAAAGCAGCANTNAGCGCANTGTA 2261
|||||:::|||||:::|||||:::|||||
804 AspLysAlaLeuAsnSerPheAsnProThr... 813
2262 NCTNNCCATTAACGNTNTTAANCTCNCNGGCTGCTCNCACNTA 2311
813 ..... 813
2312 GCAATCTTAGTCAATGCGCATACAGCTTATACAGTACGCAACGCC 2361
|||||
814 ..AsnLeu..... 815
2362 ACCCAAAACGGACCTTACCTCGTGGGCGCATGCCCAACAACATTAA 2411
:::|||||:::|||||:::|||||:::|||||
816 ....ArgGlyAsnValAsnLeuThrGluSerAlaAsnPheValLeu 830
2412 TCAAGCCACATTAAACGCAACNCATCGGTTGGGCAATGCTTATTAA 2461
:::|||||:::|||||:::|||||:::|||||
830 yLysAlaAsnLeuPheGlyThrIleGlnSerArgLysSerGlnVal 847
2462 ACTTAAGCAACAACGGCGCAACAAAGCGAGTGTGACGCTTCCGACA 2511
:::|||||:::|||||
847 rGLeuThrGluAsn..... 851
2512 GCTAAGCAACAGTAAAGCATTCGCGACATCAAGGCAATGTCTCCCTAG 2561
851 ..... 851
2562 CGATTAAGCAATTCATTTGAACAACGCCGCTTACGGACAACATCA 2611
851 ..... 851
2612 GCGGCGCAAGANCAAGATTACACTTAAGACAGCAATGACAGCTG 2661
|||||:::|||||
852 .....SerHisThrPheLeu 856
2662 CGGTGACGACGGAATTAGCAATTTAACTTGACACGCGCACCATATAC 2711
:::|||||:::|||||:::|||||:::|||||
857 ThrGlyAsnSerAspValHisGlnLeuAspLeuAlaAsnGlyHisIle 873
2712 ACTCAATTCGCGCTATCGCCACGAGTGTGACGCGCGCAACCGGCA 2761
|||||:::|||||:::|||||
873 sLeuAsnSerAlaAspAsnSerAsn..... 881
2762 TGTCAACACACGCGCGCGCGCTGCGCGCTTCCCTATTATCCGTTACA 2811
:::|||||
882 .....AsnValThr 884
2812 CGGCAACTTCGTAGTAATCCGTTTCAACAGCTGACGTAAGCGCAA 2861
:::|||||:::|||||:::|||||:::|||||
885 .....LysTyrAsnThrLeuThrValAsn...Se 893
2862 ATTGAACNGTCAAGAACATTCGCTTATGTGGAACCTCTTGCGCTACC 2911
:::|||||:::|||||:::|||||:::|||||
893 rLeuSerGlyAsnGlySerPheTyrThrLeuThrAspLeuSerAsnLys 910

2912 GAAGCACAATTTGACCTGGCGGAAAGTCCAGAGNACTTACACCTTG 2961
:::|||||:::|||||:::|||||:::|||||
910 InGlyAspLysValValValThrLysSerAlaThrGlyAsnPheThrLeu 926
2962 GCGGTCAACATTAACCGGCAACGAGACCCGTAAAGCTCATCAATTTGAC 3011
|||:::|||||:::|||||:::|||||
927 GlnValAlaAspLysThrGlyGluPro...AsnHisAsnIleThrLe 942
3012 AGTGAAGGGAAGACAAACACCGCTGCGCAAAACCTTAATTTACCC 3061
:::|||||:::|||||:::|||||:::|||||
942 uPheAspAla...SerLysAlaGlnArgAspHisLeuAsnValSerL 957
3062 TGGCAAAAGCAACAGTACCGCGCGCGCTGCGCTTACCACTCATCCGC 3111
|||:::|||||:::|||||:::|||||:::|||||
957 euValGlyAsnThrValAspLeuGlyAlaThrPlyTyrLysLeuArg 973
3112 AAGACGCGCAGTTCCTGATTAATTCGCGTCAAAAGACA..... 3153
:::|||||:::|||||:::|||||:::|||||
974 ValAsnGlyArgTyrAspLeuTyrAsnProGluValGluLysArgAsn 990
3153 ..... 3153
990 nThrValAspThrThrAsnIleThrThrProAsnAsnIleGlnAlaAsp 1007
3154 .....GAGCTTCCGACAAACTCGGCAAG..... 3177
1007 alProSerValProSerAsnAsnGluIleAlaArgValAspGluAla 1023
3177 ..... 3177
1024 ProValProProProAlaProAlaThrProSerGluThrGluThrVal 1040
3178 .GCAGAACCAAAACAGCGCGGAAAAA.....GACACGCGCAAA 3217
|||||:::|||||:::|||||
1040 lAlaGluAsnSerLysGlnSerLysThrValGluLysAsnGluAla 1057
3218 GCGTTGACGCGCTGATTCGCGCGCGCGCATGCGCGCGAA..... 3258
:::|||||:::|||||:::|||||:::|||||
1057 sPalaThrGluThrAlaGlnAsnArgGluValAlaLysGluAlaLys 1073
3259 .....AAGACAGAAAGCGGTGCGGACGAGCGGCGCG 3287
:::|||||:::|||||:::|||||:::|||||
1074 SerAsnValLysAlaAsnThrGlnThrAsnGluValAlaGln..... 1087
3288 GCGGACGCGCGGAAATGTCGCAATTATGACGCGAG..... 3327
:::|||||:::|||||:::|||||
1088 ...SerGlySerGluThrLysGluThrGlnThrThrGluThrLysGlu 1103
3328 .....GAAGGAAAAAAGCGGTGCGGACGAGCGGAAAGAC 3360
|||||:::|||||:::|||||:::|||||
1103 hrAlaThrValGluLysGluLysAlaLysValGluThrGluLysThr 1119
3361 AGCGCNTTGGCGAAA.....CAGCGGCAACGCGGAAACCGCGCGGN 3401
:::|||||:::|||||:::|||||:::|||||
1120 GlnGluValProLysValThrSerGlnValSerProLysGlnGluLys 1136
3402 TACACGCGCTTCCCGCGCGCGCGCGCGCGCGCGGAAATTTGCGG... 3447
|||:::|||||:::|||||:::|||||:::|||||
1136 rGluThrValGlnProGlnAlaGluProAlaArgGluAsnAspProThr 1153
3448 .....CAACGCGACGCCCA..... 3462
:::|||||:::|||||
1153 alaAsnIleLysGluProGlnSerGlnThrAsnThrAlaAspThrGlu 1169
3463 .....CCGCAACT..... 3471
1170 GlnProAlaLysGluThrSerSerAsnValGluLysProValThrGlu 1186
3471 ..... 3471
1186 rThrThrValAsnThrGlyAsnSerValValGluAsnProGluAsnThr 1203
```

```

3472 .....CAACC.....CAACG 3483
1203 hrProalaThrThrGlnProThrValAsnSerGluSerAsnLysPro 1219
3484 CAGGC..... 3489
1220 LysAsnArgHisAlaArgSerValArgSerValProHisAsnValGluPr 1236
3490 .....GACCTGA 3496
1236 oAlaThrThrSerSerAsnAspArgSerThrValAlaLeuGlyAspLeuT 1253
3497 TNAGCGGTATATGCCAATAGCGGTGAGTGAATTTCCGCC..... 3537
1253 hrSerThrAsnThrAsnAlaValLeuSerAspAlaArgAlaLysAlaGln 1269
3538 .....ACGCTCAACAGCGTTTCCGCTACAGAGACGAATGGACCGCT 3581
1270 PheValAlaLeuAsnValGlyLysAlaValSerGlnHisIleSerGlnLe 1286
3582 G...TTGGCGAAGACCGCGCAACGCGTTTGGACAGCAGCAGCCG 3628
1286 uGluMetAsnAsnGluGlyGlnIleArgAsnValIlePheValSerThrSe 1303
3629 ACACCAACACACTACGCTGCGAAGATTTCCGCGCTACCGCAACCAAC 3678
1303 etAsnLysAsnIleYrSerSerGlnIleArgArgPheSerSerLysSer 1319
3679 GACCTGCGCAATCGGTATGCAAAAAACCTCGCAGCGG...CGCCT 3725
1320 ThrGlnThrLeuGlyIlePheValPheValIleSerAsnAsnValGlnLe 1336
3726 CGGCATCTGTTTTCGCAACCGGACGCGCAAAACANTTCGACAGCGCA 3775
1336 uGlyValAlaPheThrValArgAsnSerAsnAsnPheAspLysAlaT 1353
3776 TCGGCACTCGGACGCGCTTCCCGCGCGCGCTTTTGGGCAATACGCG 3825
1353 hrSerLysAsnThrLeuAlaGlnValAsnPheThrSerLysIleThrAla 1369
3826 ATGGGAGGTTCGACATCGGCATACGCGCGCGGTTCATGACGCGG 3875
1370 AspAsnHisThrPheLeuGlyIleAspLeuGlyThrGlyLysPheGlnLe 1386
3876 CANTCTNTCAGACGCGCATCGGCAAAATCCGCGCGCGCTGCTCAT 3925
1386 IlyLeuGlnIleThrAsnHisAsnAlaLysPheAlaArgHisThrAlaGlnP 1403
3926 ACGGCATTACGCGCATACCGCGCGCTTGGCGGATTCGCGCATCGCA 3975
1403 heGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIleThr 1419
3976 CGGTACATCGCGCGCAACGCGGTATTCGCAAAAGCGGATTAACCGCA 4025
1420 ProIleValGlyAlaArgIleSerIleThrLeuSerAsnAlaAspPheAlaLe 1436
4026 CGAAAACGTCATATCGCCACCGCGCTTGGTCAACCGTACCGG 4075
1436 uAspGlnAlaArgIleLysValAsnProIleSerValLysThrAlaPheAla 1453
4076 CGGCGATTAGGCGATTTATTCATTCAAAACGCGCGCAACACATNTCCATC 4125
1453 IlyLeuValAspLeuSerIleThrIleHisLeuGlyGlu...PheSerVal 1468
4126 ACNCGTTATTNAGCCTGCTCCTATACCGATGCGCGTTCGGGCAAGTCG 4175
1469 ThrProIleLeuSerAlaArgIle...AspAlaAsnGlnIleSerGlyLys 1484
4176 AACACGCGTCATATACCGCGTATTCGCGATTCGCGCAAAACCGCA 4225
1484 sIleAsnValAsnGlyIleThrAspPheAlaLysValGluAsnGlnGlnG 1501
4226 GTGCGGAATGGGCGTAAACGCCGAATCAAAAGTTTCACGCTGTCNTC 4275

```

```

1501 IntyAsnAlaGlyLeuLysLeuLysThrHisAsnValLysLeuSerLeu 1517
4276 CACGCTGCCCGCCGCAAGNCCGCACTGAGACGCGCAACACACGCGGG 4325
1518 IleGlyGlyLeuThrLysAlaLysGlnAlaGlnLysGlnLysThrAlaG 1534
4326 CATCAATATTAGGCTAC 4341
1534 uLeuLysLeuSerPhe 1539

seq_name: SwissProt_40:IGA0_HAEIN
seq_documentation_block:
ID IGA0_HAEIN STANDARD; PRT; 1694 AA.
AC P44969;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA OR IGA1 OR HI0990.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC SRRAIN-SEROTYPE D;
RA Wright A., Fishman Y., Tai F., Plaut A.G.;
RL Submitted (MAY-1991) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SRRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ufferback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; X59800; ?; NOT_ANNOTATED_CDS.
CC EMBL; U32779; AAC22651.1; -
CC MEROPS; S06.001; -
CC TIGR; HI0990; -
CC InterPro; IPR000710; IGA_S6.
CC Pfam; PF02395; IGA1; 1.

```

DR PRINTS; PRO0921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal;

FT	SIGNAL	1	25	POTENTIAL
FT	CHAIN	26	1014	IMMUNOGLOBULIN A1 PROTEASE
FT	PROPER	1015	1694	HELPER PEPTIDE (POTENTIAL)
FT	ACT_SITE	288	288	PROBABLE.
FT	CONFLICT	253	254	EN -> GV (IN REF. 1).
FT	CONFLICT	272	272	G -> A (IN REF. 1).
FT	CONFLICT	464	464	G -> E (IN REF. 1).
FT	CONFLICT	866	866	S -> T (IN REF. 1).
FT	CONFLICT	1036	1036	A -> D (IN REF. 1).
FT	CONFLICT	1074	1074	A -> G (IN REF. 1).
FT	CONFLICT	1421	1421	A -> T (IN REF. 1).
FT	CONFLICT	1545	1545	H -> G (IN REF. 1).
SO	SEQUENCE	1694 AA:	185539 MM:	C52427013F03178C CR664

```
alignment_scores:      Length: 1873  
    Quality: 1519.00  
    Ratio: 1.653     Gaps: 54  
Percent Similarity: 49.066 Percent Identity: 24.826
```

```
alignment_block;
```

US-09-303-518D-651 X IGA0_HAEIN . .

Align seg 1/1 to: IGA0_HAEIN from: 1 to: 1694

```

64 GCGTTCGGCGCGTCTACTTATGACCAATAGCGCTGCGTGGCAATGTCG 113
   :::::|||||:  :::::|||||:  :::::|||||:  :::::|||||:  :::::
5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrP 21
   :::::|||||:  :::::|||||:  :::::|||||:  :::::|||||:  :::::
114 CCGACCTTGGGGGGGACACACTTATTCGGCATCCAACTACCAATACTATC 163
   |  ::  |||||:  ::  |||||:  ::  |||||:  ::  |||||:  ::  ||
21 GTCATCTGCTuAlaAlaLeuValAlaGAspAspValAlaSPtyGlnIlePheA 38
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
164 GGCATCTTGGCGAAATATTAAGCAAGTTTGGCGCGGGGCGAAAGATAT 213
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 tGAspPheAlaLeuAsnLysSerLysTyrPheSerValGlyAlaThrAspVal 54
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
214 GAGGTNTACAAACAAAAAGGGAGACTTGGTCGCGCAATCAATGACAAAGC 263
   |||||  :::::|||||:  :::::|||||:  :::::|||||:  :::::
55 GluValAlaGAspLysAsnAsnHisSerLeuGlnValLeuProAsnGln 71
   |||||  :::::|||||:  :::::|||||:  :::::|||||:  :::::
264 C...CGGATGATGATTTTCTGTGGTGTGC...CGTACGGCGGTGGCG 307
   |  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
71 YIleGrometIleAspPheSerValValAlaSPValAlaSPValGlyAla 88
   ::  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
308 CATTCGTGGCGGCATATATATTTGTCGACCGCGGCACAT...AACGGCGGC 354
   ::  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88 hIleuIleAsnProGlnTyrAlaValAlaGlyValLysHisValSerAsnGly 104
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
355 TATTAACAACGGTGAATTTGGT.....GCGGAAGGAGAGAA 389
   :::::||||  :::::||||  :::::||||  :::::||||  :::::
105 ValSerGlnIleuHisPheGlnAsnLeuAsnGlnAsnMetAsnAsnGlnAs 121
   |||||:||||  |||||:||||  |||||:||||  |||||:||||  |||||:
390 TCCCGATGACGACCGCT.....TTTCTTACCAATTG 421
   |  |||||:||||  |||||:||||  |||||:||||  |||||:||||  |||||:
121 nAlaLysSerHisAlaGAspValSerSerGlnGlnAsnArgTyrPheSerV 138
   |||||:||||  |||||:||||  |||||:||||  |||||:||||  |||||:
422 TCGAAAGAAATTAATAT.....AAG 441
   |||||:||||  |||||:||||  |||||:||||  |||||:||||  |||||:
138 alGlnLysnGlnIleuProThrIleLysLeuAsnGlnLysAlaValThrThe 154
   |||||:||||  |||||:||||  |||||:||||  |||||:||||  |||||:
442 CCGTGAATTCACACCGCTTACACGCGCATTAACATATGCGCGCTTTGCA 471
   |||||:||||  |||||:||||  |||||:||||  |||||:||||  |||||:
155 GluAspGlnThrGlnLysAlaGArgLysAspTyrTyrMetProArgLeuAs 171
   |||||:||||  |||||:||||  |||||:||||  |||||:||||  |||||:
492 TAAATTTGCACAGATGACGAAAGCTGTGGAA...ATGACGAGTACAGATGA 538
   |||||:||||  |||||:||||  |||||:||||  |||||:||||  |||||:
171 PysPheValThrGluValAlaProIleGlnAlaSerThrAlaSerGera 188
   |||||:||||  |||||:||||  |||||:||||  |||||:||||  |||||:
539 GGGGGGATCCTATTCCGATTAAGAAAAATGTCCGAGCGTGTCCGATC 588

```

[illegible]

```
441 hTTrpLysValHisAsnProLysSerAspArgLeuAlaLysIleGlyLys 457
1366 GGCACCGCTGACGCTTCACAGCCAAAGGGGAAAACCAAGGCTGATCAGCCT 1415
458 GlyThrLeuIleValGluGlyLysGlyLysGlySerLeuLysVal 474
1416 GGGCGACGGTACAGTCATTTGGATCCAGGACGACGACGATAAAGCAAAA 1465
474 IGlyAspGlyThrValIleLeuLysGlnGlnIleAspAlaAsnAsnLysVal 491
1466 AACAAACCTTTAGTGAATCCGCTTGNTCAGCGCAGGGGTACGGTCCAA 1515
491 aLysIleAspSerGlnValIGlyIleValSerGlyArgSerThrValVal 507
1516 CTGAATGCGCATATCATCAGTTCACCCCGACCAACCTATTTGGCTTTCG 1565
508 LeuAsnAspAspLysGlnValAspProAsnSerIleTyrPheGlyPheArg 524
1566 CGGCGGACGTTTGGATTTAAACGGGCATTCGCTTCACCGCTATTC 1615
524 gGlyGlyArgLeuAspAlaAsnGlyAsnAsnLeuThrPheGluHisIleVal 541
1616 AAATATACCGATGAAAGGGCGATGATTCGNCATCATATATCCCAACAACA 1665
541 rGAsnIleAspAspGlyAlaArgLeuValAsnHisAsnThrSerLysThr 557
1666 TCACCGCTTACCATTCACAGGATGAATGATATTCACACACGAGT 1710
558 SerThrValThrIleThrGlyLysLeuIleThrAspProAsnThrIle 574
1710 1710
574 eThrProTyrAsnIleAspAlaProAspGluAspAsnProTyrAlaPheA 591
1710 1710
591 rGArgIleLysAspGlyGlyGlnLeuTyrLeuAsnLeuGluAsnTyrThr 607
1710 1710
608 TyrTyrAlaLeuArgLysGlyAlaSerThrArgSerGluLeuProLysAs 624
1711 1719
624 nSerGlyLysAsnGluAsnTyrLeuTyrMetGlyLysThrSerAspG 641
1720 1752
641 LuAlaLysArgAsnValMetAsnHisIleAsnAsnGluArgMetAsnGly 657
1753 TACAACGGTGGTGGCGAGAAAGATGACCAAAACGAGGCGCGCT 1802
658 PheAsnGlyTyrPheGlyGluGluGlu...GlyLysAsnAsnGlyAsnLe 673
1803 CAACCTGTTTACCGCCCGCGGACGACCGCACCCGCTGCTTCG 1852
673 uAsnValThrPheLysGlyLysSerGlnGlnAsnArgPheLeuLeuThrG 690
1853 GCGGAACAAATTTAAACGCAACATCAGCAAAACAAACGCAAACTGTT 1902
690 LysGlyThrAsnLeuAsnGlyLysPheLysValIGlyLysGlyThrLeuPhe 706
1903 TTCAGCGGACGACGACCGCGACGCTTACATATTTAGGAGCGGCTG 1952
707 LeuSerGlyArgProThrProHisIleAlaArgAspIleAlaGlyIleSer 723
1953 GTCAAAATGAGAGT...ATCCACACAGCAAAATCGTGGGACA 1996
723 rThrLysLysAspGlnHisPheAlaGluAsnAsnGluValValIGluVal 740
1997 ACGAGCTGATNACCGCAGCTTTAAAGCGAAATTTCCATATTCAGGCG 2046
740 spAspTrpIleAsnArgAsnPheLysAlaThrAsnIleAsnValThrAsn 756
2047 GGCAGCGCGGTGATTTCC...CGCATGTTGCCAAATGCAAGCGCATTC 2093
757 AsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIleThrSerAsnIle 773
2094 NCATTTGAGCATATCAGCCCAAGCAGATTTTGGTGGCGACCGCATCAAA 2143
773 eThrAlaSerAspAsnAlaLysValHisIleGlyTyr...LysAlaG 788
2144 GGCATACAAATCTGTACAGCTTCGAGCTGACAGGCTGACAAATTTGTC 2193
788 LysThrValLysValArgSerAspTyrThrGlyThrValThrCysThr 804
2194 GAANAANNCATTTACGACGATTAAGATGATGCTTCATGCTAAGACNGA 2243
805 ThrAspLysLeuSer...AspLysAlaLeuAsnSerPheAsnAlaThrAs 820
2244 CMTNAGCGGCGAMTGNAGCTNNMCNATNACGNTNNTNAAACCTCNCNG 2293
820 nValSer... 822
2294 GGCNTGCNMCATNANGCAATCTTAGTCAAAATGGCGATACAGTTAT 2343
822 822
2344 ACAGTACAGCCAAACGCCACCAAAACGGCAACCTTAGCTGTGGGCAA 2393
823 830
2394 TGCCCAAGCAACATTTAATCAAGCCACATTTAAACGGCAACNCTCGGNT 2443
830 nAlaAsnPheValLeuGlyLysAlaAsnLeuPheGlyThrIleSerGlyT 847
2444 CGGCAATGCTTCATTTAATCTAAGCAACACCGCGCACAAAACGCACT 2493
847 hrGlyAsnSerGlnValArgLeuThrGluAsn... 857
2494 CTGACGCTTTCGCAACGCTAAGCAACGTAAGCATTCGCACTCAA 2543
857 857
2544 CGGCATGTCTCCCTACCGCATTAAGCAGATTTTCATTTGAAAACAGCC 2593
857 857
2594 GCTTACCGGACAACACTCAGCGGACGACGAGGANNACAGATTACCTTAAA 2643
857 857
2644 GACAGCAATGACGCTGCCGTCAGGACGAGCAATTAAGCAATTTAAACCT 2693
858 ...SerHisTrpHisLeuThrGlyAspSerAsnValAsnGlnLeuAsnLe 873
2694 TGCAACAGCCACCATTTACACTCAATTCGCGCTATTCGCGACAGATGCTGAC 2743
873 uAspLysGlyHisIleHisLeuAsnAla...GlnAsnAspAlaAsnL 888
2744 GCGCGCAACCGCGCAGNGTGTGACACAGCGCGCGCGCTTCGCGCGCT 2793
888 yValThrThr... 891
2794 TCCCTATTATCCGTTACACCGCCAACTTCGTAGATCCCGTTTCACAC 2843
892 894
2844 GCTGACGATTAAGCAACATTTGACNGTCAAGGACATTCGCTTATGT 2893
894 rLeuThrValAsn...SerLeuSerGlyAsnGlySerPheTyrTyrLeuT 910
2894 CGGAACCTCTTGGCTACGAGGACAAATTTAGATGGCGGAAAGTTCC 2943
910 hrAspLeuSerAsnLysGlnGlyAspLysValValValThrLysSerAla 926
```



```

1523 AspAsnHisTrpTyrLeuGlyIleAspLeuGlyTyrGlyLysPheGlnSe 1539
3876 CANTCTNFCAGACGCGATCGAGGCAAAATCCGCCCGCGTCGCAT 3925
1539 rAnLeuLysTrpAsnHisAsnAlaLysPheAlaArgHisTrpAlaGlnP 1556
3926 ACAGCATTCAGACGACGATACCGCGCGGCTTGGCGGATTCGCATCGAA 3975
1556 heLLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPhgGlyLeuThr 1572
3976 CGGTACATCGCGCAACGGCGCTATTCCGTCCAAAAAGCGGATACCGGTA 4025
1573 PrIleValGlyValArgTyrSerTyrLeuSerAsnAlaAsnPhaLe 1589
4026 CGAAACGTCATATCGCCACCCCGCTCTGCGTTCACCCNTAACCGNG 4075
1589 uAlaLysAspArgIleLysValAsnProIleSerValLysTrpAlaPheA 1606
4076 CGGCGATTAAAGCAGATTATGATTCACAAACGCGGCACATNTCCATC 4125
1606 laGlnValAspLeuSerTyrTrpHisLeuGlyGlu...PheSerVal 1621
4126 ACCGCTATTTCAGCCTGTCTAT...ACCGATCGCGCTTCGGCGCAAGT 4172
1622 ThrProIleLeuSerAlaArgTyrAspTrpHisGlnGlySerGlyLys 1638
4173 CGGACACGCGTCATATCCGCGTATGCGTCAGATTTCGGCAAAACCC 4222
1638 e.....AsnValAsnGlnTyrAspPheAlaTyrAsnValGluAsnGln 1653
4223 GCAGTCCGGAATGGCGGTAAAGCCGGAATGCAAAAGTTTCACGCTGCC 4272
1653 InGlnTyrAsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSer 1669
4273 NTCCACGCTCGCGCGCGCAAGNCCGCAATGCGACAGCCACACAGCGC 4322
1670 LeuIleGlyGlyLeuThrLysAlaLysGlnAlaGluLysGlnLysThrAl 1686
4323 GGGCATCAATTAAGGCTAC 4341
1686 aGluLeuLysLeuSerPhe 1692
seq_name: SwissProt_40:IGA2_HAEIN
seq_documentation_block:
ID IGA2_HAEIN STANDARD; PRT; 1702 AA.
AC P45384;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=127;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK715 / SEROTYPE B;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
certain Pro-I-Xaa bonds in the hinge region. No small molecule
substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
SPACE. AND THE CARBOX-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE

```

```

CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M87489; AAA24966.1; -.
CC DR MEROPS: S06.001; -.
CC DR InterPro: IPR000710; IGA_S6.
CC DR Pfam: PF02395; IGA1; 1.
CC DR PRINTS: PRO00921; IGASERPRASE.
CC KW Hydrolyase; Serine protease; Transmembrane; Zymogen; Repeat; Signal.
CC FT SIGNAL 1 25
CC FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
CC FT PROPEP 1015 1702 HELPER PEPTIDE (POTENTIAL).
CC FT ACT_SITE 288 288 PROBABLE.
CC FT DOMAIN 1109 1124 2 X 8 AA TANDEM REPEATS OF A-K-V-E-K-E-E-
CC FT REPEAT 1109 1116 K.
CC FT REPEAT 1117 1124 1.
CC FT REPEAT 1117 1124 2.
CC SQ SEQUENCE 1702 AA; 186539 MW; 860F70D267807A6 CRC64;

alignment_scores:
Quality: 1517.00 Length: 1881
Ratio: 1.651 Gaps: 53
Percent Similarity: 48.857 Percent Identity: 24.615

alignment_block:
US-09-303-518D-651 x IGA2_HAEIN ..
Align seg 1/1 to: IGA2_HAEIN from: 1 to: 1702

64 GCGTTCGCGCGCTGCTTACGATATGCGTCGCGTTCGATTCG 113
1 LysPheLysLeuAsnPhelIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
114 CCAGCTTGGGCGGACACACTTATTCGGCATCACTACCAATATCATC 163
1 LysPheLysLeuAsnPhelIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
21 oTyrThrGluAlaIleAlaLeuValArgAspAspValAspTyrGlnIlePheA 38
164 GCGACTTTCGCGGAATTAAGGCAAGTTTCAGTCGCGGCGGAAGATAT 213
1 LysPheLysLeuAsnPhelIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
38 rGAspPheAlaGluAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54
214 GAGGTATACAAACAAAAAGGAGGAGTGTGCGGCAATCATGACAAAGC 263
1 LysPheLysLeuAsnPhelIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
55 GluValArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnG 71
264 C...CCGATGATTCATTTTCTGTGTGTG...CGTACGGCGGTGCGG 307
1 LysPheLysLeuAsnPhelIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
71 yIlePrometIleAspPheSerValValAspValAspLysArgIleAla 88
308 CATTGGTGGCGGATCATATATTTGTGAGCGTGACAT...AACGGCGC 354
1 LysPheLysLeuAsnPhelIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
355 TATAACAACGTGATTTTGGT...GCGGAAGGAAGMA 389
1 LysPheLysLeuAsnPhelIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnGlyAs 121
390 TCCGATCAGCACCGT...TTTCTTACCAATTG 421
1 LysPheLysLeuAsnPhelIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
121 nAspLysSerHisArgAspValSerSerGluGluAsnArgTyrPheSerV 138
422 TGAAGAAGAAATTAATTAT...AAG 441
1 LysPheLysLeuAsnPhelIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

```



```

138 aLgLUyAsnGluTyrProThrLysLeuAsnGlyLysAlaValThrThr 154
442 CCTGCAATTCACACCTTACACGGGCGATATACATATGCGCGCTTGCA 491
155 GUAAspGlnThrGlnLysArgGlnAspTyrTrpMetProAlaGlnAs 171
492 TAAATTTGTCACAGATGACGAACTGTGCA...ATGACAGTGCATGA 538
171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerSera 188
539 GGGGAATACCTATTCGATTAAGAAAATATCCGAGCGCTGCCATC 588
188 sPAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204
589 GGGTCAGACACCACTATGGCGTTATGATGACAAACGCGCGATTT 638
205 GlySerGlySerGlnPheIleTyr.....LysLysGlyAsp.. 216
639 ATCTACTCCGGCGCATGTAAATGGCGCAATACATATGACGGGTT 688
217 .AsnTyrSer.....LeuIleLeuAsnAsnHisGluValGly. 228
689 GGGGAATTAAT.....GGCGTANTT 708
229 ..GlyAsnAsnLeuLysLeuValGlyAspAlaTyrThrTyrGlyIleAla 244
709 AGTTGACGCGCGATGTGCGCATCCACAGAC.....TATGG 746
245 GlyThrProTyrLysValAsnHisGlnAsnGlyLeuIleGlyPheGly 261
747 C.....C 748
261 yAsnSerLysGluGlnHisSerAspProLysGlyIleLeuSerGlnAsp 278
749 CTATGCCATTCGACGTCCGCGACGCGACGCGGTGCGCAATGTTATT 798
278 rOleuThrAsnLysAlaValLeuGlnLysAspSerGlySerProLeuPheVal 294
799 TATGCAAAACAAACAAATAAATGGCGTCAACGAGTTTACAACCGG 848
295 TyrAspAlaGlyLysGlyLysTyrLeuPheLeuGly.....Se 307
849 CTACCCATTATCCGCGACGAGAAACGGTTTCACGTGATACCAAGATT 898
307 rTyrAspPheThrPalaGlyTyrAsn.....LysLysSerT 319
899 GGTTCACGATGACATTTACAGAGCGCATACACATACCGCTNTTTGAA 948
319 rPAlaGlnLysTyrAsnLysTyrLysProGlu..... 328
949 CCGCGCATTAACGCGATTTTCCTTTACATCCAAACAAACGCGTACG.. 996
329 .....PheAlaLysThrValLeuAspLysAspThrAl 339
997 .GGTACGGTAAACAGAAACCAAGAAAGTTTC...AATCCAAAGCTTA 1042
339 aGlySerLeuThrGlySerAsnThrGlnTyrAsnThrAsnProThrGlyL 356
1043 AAGTACACAGCTCCGACTGTTTGACGCAATCTTGAAT..... 1080
356 yThrSerValIleSerAsnGlySerLysSerLeuAsnValAspLeuPhe 372
1081 .....GAACGTATTAAGAAACCACTTACGCGGCAAGGGGTGT 1118
373 AspSerSerGlnAspThrAspSerLys..... 381
1119 TAATCAGTACGCTCCAGGTTAAACAAACGCGTAAACCTTTCTTTATCG 1168
382 .....LysAsnAsnHisGlyLysSerValThrLeu... 391
1169 ATTACGCAAGCGCAAACTCATCTTATCAAAACATCAACCAAGCGCG 1218
392 ..ArgGlySerGlyThrLeuThrLeuAsnAsnHisLeuAspGlnGlyAla 407

1219 GCGCGTTTGATTTTGAAGTCATTTTACGGTC...TCGCGTGAACAA 1265
408 GlyGlyLeuPhePheGluGlnGlyAspTyrGlnValLysGlyThrSerAsp 424
1266 CGAAGCGTGGCAAGCGCGCGCTCATATCACTGAAGACATACCGTTA 1315
424 rThrThrThrPlyGlyAlaGlyAlaSerValAlaAspGlyThrValT 441
1316 CTGGAAAGTAAACGGCGTGGCAACGACCGCTGTCCAAATCGGCAAA 1365
441 hTrpLysValHisAsnProLysSerAspArgLeuAlaLysIleGlyLys 457
1366 GGCACGCTGCACGTTCACAGCCAAAGGAGAAACCGCTGATTCGCGCTT 1415
458 GlyThrLeuIleValGlnGlyLysGlyGlnAsnLysGlySerLeuLysVa 474
1416 GGGCGACGGTACAGTCATTTGGATCAGACGACGACGATTAAGGCAAA 1465
474 lGlyAspGlyThrValIleLeuLysGlnGlnAlaAspAlaAsnLysV 491
1466 AACAGCCTTTAGTGAATGCGCTTGNTACGCGCGAGGGGTACGGTGA 1515
491 aLysAlaPheSerGlnValGlyIleValSerGlyArgSerThrValVal 507
1516 CTGAATGCCGATATACGTTCAACCGCGCATTCGTTGCTTCCACCGTATC 1565
508 LeuAsnAspAspLysGlnValAspProAsnSerIleTyrPheLysPhe 524
1566 GCGCGACGTTTGGATTAACCGCGCATTCGTTGCTTCCACCGTATC 1615
524 gGlyGlyArgLeuAspAlaAsnGlnLysAsnLeuThrPheGlnHisIleA 541
1616 AAATACCGTGAAGGCGCATGTCNCATATATATGCGCACACAAACA 1665
541 rGAsnHisAspAspGlyAlaArgLeuValAsnHisAsnThrSerLysThr 557
1666 TTCACCGTTCATTAACGAGGAAAGTAAAGTATACCAACCGGAGT... 1710
558 SerThrValThrIleThrGlyGlnSerLeuIleThrAspProAsnThrI 574
1710 ..... 1710
574 eThrProTyrAsnIleAspAlaProAspGlnAspAsnProTyrAlaPheA 591
1710 ..... 1710
591 rGArgIleLysAspGlyGlyGlnLeuLysLeuAsnLeuGlnAsnTyrThr 607
1710 ..... 1710
608 TyrTyrAlaLeuArgLysGlyAlaSerThrArgSerGlnLeuProLysAs 624
1711 .....GGTAGAAT..... 1719
624 nSerGlyGlnSerAsnGlnAsnThrPleuTyrMetGlyLysThrSerAspG 641
1720 .....ATCAATAGACTTAATTACGCAAGAAATATGCG 1752
641 lAlaLysArgAsnValIleMetAsnHisIleAsnAsnGlnIleArgMetAsnGly 657
1753 TACACGCGTGGTTGGCGAGAAAGTACGACCAAAACGACGCGGGCT 1802
658 PheAsnGlyTyrPheGlyGlnGlnGlu...GlyLysAsnAsnGlyAsnLe 673
1803 CAACCTTGTTTTACACCGCGCGCAGAGAAGCGCACCGCTGCTTCCG 1852
673 uAsnValThrPheLysGlyLysSerGlnGlnAsnArgPheLeuThrG 690
1853 GCGGAACAATTTAAACGCGCAACATCACGCAAAACGCAAACTGTT 1902
690 lGlyThrAsnLeuAsnGlyAspLeuLysValGlnLysGlyThrLeuPhe 706

```

```
1903 TTCAGCGGAGACGACGACGCGCTTACATTCATTGAGAGCGGGTG 1952
      ::::::::::::::::::::
707 LeuSerGlyArgProThrProHisAlaArgPheIleAlaGlyIleSerSe 723
      ::::::::::::::::::::
1953 GTCAAAAATGGAAGGT.....ATCCACAGAGAAATCGTGTGGACA 1996
      ::::::::::::::::::::
723 rThrLysLysAspGlnHisPheAlaGlnAsnGlnValValAlaGlnA 740
      ::::::::::::::::::::
1997 ACGACGTGTCNACGCGCATTTAAAGCGGAAATTCATATTCAGGCG 2046
      ::::::::::::::::::::
740 spAspTrpIleAsnArgAsnPheLysAlaThrAsnIleAsnValThrAsn 756
      ::::::::::::::::::::
2047 GGGCAGCGGTGATTTC...CGCAATGTGTCCAAAGTGAAGCGGATTC 2093
      ::::::::::::::::::::
757 AsnAlaThrLeuTyrSerGlyArgAsnValAlaAsnIleThrSerAsnI 773
      ::::::::::::::::::::
2094 NCATTGTGACATCAGCCCAAGCAGTTTGTGTGCGACCGCATCAAA 2143
      ::::::::::::::::::::
773 eThrAlaSerAspAsnAlaLysValHisIleGlyTyr.....LysAlaG 788
      ::::::::::::::::::::
2144 GCCATCAATCTGTACAGCTGCGACTGGACNGCTGTGACAAATGTGTC 2193
      ::::::::::::::::::::
788 LysPheValCysValArgSerAspTyrThrGlyTyrValThrCysTrp 804
      ::::::::::::::::::::
2194 GAANAANCAATTACGCGACGATAAGTGTGCTCATTTGACTAAGACNGA 2243
      ::::::::::::::::::::
805 ThrAspLysLeuSer...AspLysAlaLeuAsnSerPheAsnAlaThrAs 820
      ::::::::::::::::::::
2244 CATTNAGCGCATGTNAGCTNCCNATNACGTTNNTTNAANCTCNCNG 2293
      ::::::::::::::::::::
820 nValSer..... 822
      ::::::::::::::::::::
2294 GGCNTCCNNCATTNANGGCAATCTTAGTCAATGGCGATACGTTAT 2343
      ::::::::::::::::::::
822 ..... 822
      ::::::::::::::::::::
2344 ACAGTCAGCCACACGCGCAACCGCAACCTTAGCCTGTGGGCA 2393
      ::::::::::::::::::::
823 .....GlyAsnValAsnLeuSerGlyAs 830
      ::::::::::::::::::::
2394 TGCCCAAGACATTTAATCAAGCCACATTTAAAGCGCAACNCATCGGNT 2443
      ::::::::::::::::::::
830 nAlaAsnPheValLeuGlyLysAlaAsnLeuPheGlyThrIleSerGlyT 847
      ::::::::::::::::::::
2444 CGGGCAATGCTCATTTAATCTAAGCAACGCGCACAAAAGCGCAGT 2493
      ::::::::::::::::::::
847 hGlyLysnSerGlnValArgLeuThrGluAsn..... 857
      ::::::::::::::::::::
2494 CTGACGCTTTCGACAAAGCTAAGCAAAAGTAAAGCATTCGCGACGAA 2543
      ::::::::::::::::::::
857 ..... 857
      ::::::::::::::::::::
2544 CGGCAATGTCTCCCTAGCGGATAGGCAATTCATTGTAACACAGCC 2593
      ::::::::::::::::::::
857 ..... 857
      ::::::::::::::::::::
2594 GCTTACCGGACACTCAGCGGACGAGAGANACAGATTACACTTAAAA 2643
      ::::::::::::::::::::
857 ..... 857
      ::::::::::::::::::::
2644 GACAGGAATGAGCGCTGCGCTCAGGACGAGAAATAGCAATTTAAACCT 2693
      ::::::::::::::::::::
858 ...SerHisTrpHisLeuThrGlyAspSerAsnValAsnGlnLeuAsnIle 873
      ::::::::::::::::::::
2694 TGAACAAGCCACCATTTACACTCAATTCGCGCTAAGCGCAAGTGTGAG 2743
      ::::::::::::::::::::
873 uAspLysGlyHisIleHisLeuAsnAla.....GlnAsnAspAlaAsnL 888
      ::::::::::::::::::::
2744 GGGCGCAAAACCGGACNGTGTCAAGACACGCGCGCGCGCTGCGCGCT 2793
      ::::::::::::::::::::
888 yValAlaThr..... 891
      ::::::::::::::::::::
2794 TCCCTATTATCCGTTACACCGGCAACTTCGTAAGTAACCGTTCAACAC 2843
      ::::::::::::::::::::
      ::::::::::::::::::::
892 .....TyrAsnTh 894
      ::::::::::::::::::::
2844 GCTACGCGTAAACGCGCAATTTGAACNGTCAAGGACATTCGCGTTATCT 2893
      ::::::::::::::::::::
894 rLeuThrValAsn...SerLeuSerGlyAsnGlySerPheTyrTyrLeuT 910
      ::::::::::::::::::::
2894 CGAAGCTCTTCGCTACCGAAGCGACAAATTTGAAGTGGCGGAAAGTTC 2943
      ::::::::::::::::::::
910 hAspLeuSerAsnLysGlnGlyAspLysValValAlaThrLysSerAla 926
      ::::::::::::::::::::
2944 GAAGNACTTACACCTTGGCGGTCAACATACCGGACAGACACCGCTAAG 2993
      ::::::::::::::::::::
927 ThrLysnPheThrLeuGlnValAlaAspLysThrGlyLupProThrLys 943
      ::::::::::::::::::::
2994 CCTGATCAATTTGACGTAGTGGAGGAGGAAAGCAACAAACCCGTGCC 3043
      ::::::::::::::::::::
943 s...AsnGlnLeuThrLeuPheAspAla.....SerAsnAlaThrArg 957
      ::::::::::::::::::::
3044 AAACCTTAATTTCAACCTGCAAAACGACACGTCATGCGCGCGCTGG 3093
      ::::::::::::::::::::
957 snAsnLeuAsnValSerLeuValGlyAsnThrValAlaPheLysAlaTrp 973
      ::::::::::::::::::::
3094 CGTTACCAACTCATCCGCAAAAGCGGCGAGTTCGCTGATATTCG.. 3141
      ::::::::::::::::::::
974 LysTyrLysLeuArgAsnValAsnGlyArgTyrAspLeuTyrAsnProG 990
      ::::::::::::::::::::
3141 ..... 3141
      ::::::::::::::::::::
990 uValGlnLysArgAsnGlnThrValAspThrThrAsnIleThrThrPro 1007
      ::::::::::::::::::::
3141 ..... 3141
      ::::::::::::::::::::
1007 snAsnIleGlnAlaAspValProSerValProSerAsnAsnGlnGluIle 1023
      ::::::::::::::::::::
3141 ..... 3141
      ::::::::::::::::::::
1024 AlaArgValGluThrProValProProProAlaProAlaThrProSerG 1040
      ::::::::::::::::::::
3141 ..... 3141
      ::::::::::::::::::::
1040 uThrThrGluThrValAlaGlnAsnSerLysGlnGluSerLysThrValG 1057
      ::::::::::::::::::::
3141 ..... 3141
      ::::::::::::::::::::
1057 LuLysAsnGlnGlnAspAlaThrGluThrThrAlaGlnAsnGlyGluVal 1073
      ::::::::::::::::::::
3141 ..... 3141
      ::::::::::::::::::::
1074 AlaGlnGluAlaLysProSerValLysAlaAsnThrGlnThrAsnGlu 1090
      ::::::::::::::::::::
3142 .....GTCAAG 3148
      ::::::::::::::::::::
1090 lAlaGlnSerGlySerGluThrGlnGluThrGlnThrGluIleLysG 1107
      ::::::::::::::::::::
3149 AACAAAGACTTTCGACAAACTCGGCAAGCGCAGAGCAAAAGAAAGAG... 3195
      ::::::::::::::::::::
1107 LuThrAlaLysValGluLysGlnGluLysAlaLysValGluLysGlnGlu 1123
      ::::::::::::::::::::
3195 ..... 3195
      ::::::::::::::::::::
1124 LysAlaLysValGluLysAspGlnIleGlnGluAlaProGlnMetalase 1140
      ::::::::::::::::::::
3195 ..... 3195
      ::::::::::::::::::::
1140 rGluThrSerProLysGlnAlaLysProAlaProLysGluValSerThra 1157
      ::::::::::::::::::::
3195 ..... 3195
      ::::::::::::::::::::
1157 spThrLysValGluGluThrGlnValGlnAlaGlnProGlnThrGlnSer 1173
      ::::::::::::::::::::
3195 ..... 3195
      ::::::::::::::::::::
```

1174 ThrThrValAlaAlaAlaGluAlaThrSerProAsnSerLysProAlaGlu 1190
 3196GGGAAAAAGACGGCAAGCCTTGACCGCGTGA 3232
 1190 uGluThrGlnProSerGluLysThrAsnAlaGluProValThrProVal 1207
 3233 TTGCGCGCGGG..... 3243
 1207 aSerLysAsnGlnThrGluAsnThrThrAspGlnProThrGluArgGlu 1223
 3244 CGCGATGCCCGCGCAAGACGAAAGCGCTGCCACCGCGCCGCGGCGC 3293
 1224 LysThrAlaLysValGluThrGluLysThrGlnGluProGlnValAla 1240
 3294 AGGC..... 3297
 1240 aSerGlnAlaSerProLysGlnGluGlnSerGlnThrValGlnProGlnA 1257
 3298GGGGAATATGTC...GGCATTATGACAGCGGAGAG 3333
 1257 LaValLeuGluSerGluAsnValProThrValAsnAsnAlaGluGluVal 1273
 3334 AAAAAGCGGTGCGAGCGGATAAAGAC..... 3360
 1274 GlnAlaGlnLeuGlnThrGlnThrSerAlaThrValSerThrLysGlnPr 1290
 3361AGCGNTTGGCGAAGC 3376
 1290 cAlaProGluAsnSerIleAsnThrGlySerAlaThrAlaIleThrGluT 1307
 3377 AGCGCGAAGCGGAAACCGCGCGGNTATC..... 3405
 1307 hrAlaGluLysSerAspLysProGlnThrGluThrAlaAlaSerThrGlu 1323
 3405 3405
 1324 AspAlaSerGlnHisLysAlaAsnThrValAlaAspAsnSerValAlaAs 1340
 3406ACCGCTTCCCGCGCGCGCGCGCGCGGAGATTGC 3445
 1340 nasSerLysSerSerGluProLysSerArg...ArgArgArgSerLies 1356
 3446 CGCAACCGCAC..... 3456
 1356 ergLInProGlnGluThrSerAlaGluGluThrThrAlaAlaSerThrAsp 1372
 3457CCCAACCGCAACCTCAACCCCAAC 3482
 1373 GluThrThrIleAlaAspAsnSerLysArgSerLysProAsnArgArgSe 1389
 3483 GCAG..... 3486
 1389 rArgArgSerValArgSerGluProThrValThrAsnGlySerAspArgS 1406
 3487CGGACCTGATNAGCGGTATGCCAATAGCGGTTTG 3522
 1406 erThrValAlaLeuAlaArgAspLeuThrSerThrAsnThrAsnAlaValIle 1422
 3523 AGTGAATTTCCGCC.....ACGCTCAACAGCGTTTCCG 3557
 1423 SerAspAlaMetAlaLysAlaGlnPheValAlaLeuAsnValGlyLysAl 1439
 3558 CGTACAGACGAATTTGACCGCGT...TTTGGCGAAGACCGCGCGCAACG 3604
 1439 aValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluGlyLysIle 1456
 3605 CNGTTTGGACAGCNGCATCCGGAACCAAACTACCGTTGCCAAGAT 3654
 1456 snValITPValSerAsnThrSerMetAsnGluAsnTyrSerSerSerGln 1472
 3655 TTCCGCGCTTACCGCAACAGACCTGCGCAATCGGTATGCGAA 3704
 1473 TyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGlyTyrPaspGl 1489

3705 AACTCTGGGACGGG...CGCGTGGCATTCCTTTTGGCAACCCGA 3751
 1489 ntThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTyrValAla 1506
 3752 CCGAAACANCTTTCGACGACGCGATCGCAACCTCGACGCGCTGCCAC 3801
 1506 snSerAsnAsnPheAspLysAlaSerSerLysAsnThrIleuAlaGlnVal 1522
 3802 GCGCGCGTTTGGGCATACGGCATCGCAGGTTTCGACATCGGCATCAG 3851
 1523 AsnPheTyrSerLysTyrTyrAlaAspAsnHisTyrTyrLeuGlyIleAs 1539
 3852 CACGGCGCGGTTTTCGACGCGCGCANCTGTCACAGCGCATCGGACGA 3901
 1539 PleuGlyTyrGlyLysPheGlnSerAsnLeuLysThrAsnHisAsnAla 1556
 3902 AATTCGCGCGCGCTGTCATTCAGCATTCAGCATTCAGCATTCAGCGCGC 3951
 1556 yspHealatrGlnHisThrAlaGlnPheGlyLeuThrAlaGlyLysAlaPhe 1572
 3952 GGTTCGCGGATTCGCGCATCGACCGTACATCGCGCGCAACGCGTATT 4001
 1573 AsnLeuGlyAsnPheGlyIleThrProIleValGlyValArgTyrSerTy 1589
 4002 CGTCAAAAAAGCGATTACCGGTACGAAACGTCATTCGCCACCCCGG 4051
 1589 rLeuSerAsnAlaAsnPheAlaLeuAlaLysAspArgIleLysValAsnPr 1606
 4052 GTCCTGCTTCACCGCATCCGCGCGGATTAAGCGATTAATTCATTC 4101
 1606 rolLeuSerValLysThrAlaPheAlaGlnValAspLeuSerTyrThrTyr 1622
 4102 AATCCGCGCAACATNTCCATACNCCTTATTNAGCTGTCCAT... 4149
 1623 HisLeuGlyGlu...PheSerValThrProIleLeuSerAlaArgTyrAs 1638
 4150 .ACCGATGCGCGCTTCGCGCAAGTCCGACACGCGCATTCACCGCGTAT 4198
 1638 pThrAsnGlnGlySerGlyLysIle....AsnValAsnGlnTyrAsp 1653
 4199 TGCTCTCAGATTTCGGCAAAACCGCGAGTGGCGAATGGGCGCTAAACGCC 4248
 1653 heaLatyrAsnValGluAsnGlnGlnTyrAsnAlaGlyLeuLysLeu 1669
 4249 GAATCAAGGTTTCACGCTGTCNCATCCACGCTGCCCGCGCAAGGNC 4298
 1670 LysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThrLysAlaL 1686
 4299 GCAACTGGAAGCGCAACACGCGCGGCGATCAATTAAGCTTAC 4341
 1686 sglnAlaGluLysGlnLysThrAlaGluLeuLysLeuSerPhe 1700

seq_name: Swissprot_40:IGA_NEIGO

seq_documentation_block:
 ID IGA_NEIGO STANDARD; PRT; 1532 AA.
 AC P09790:
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE IGA-specific serine endopeptidase precursor (EC 3.4.21.72) (IGA
 DE protease).
 GN IGA.
 OS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=485;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=MS11;
 RX MEDLINE=87115823; PubMed=3027577;
 RA Pohlner J., Hatter R., Beyreuther K., Meyer T.F.,
 RT "Gene structure and extracellular secretion of Neisseria gonorrhoeae


```
353 .....Alavala 355
1058 GACTGTTGACGAATCTTTGAATGAAGTATGAAGCAAGTTTACGG 1107
      |||||
355 rgleuAlaasn.....AsnGluGlyAsp..... 362
1108 GCAGGGGGGTGTAATCAGTACCGTCCAAAGGTTAAACAAAGGTGAACCT 1157
      |||||
363 .....AlaasnAsnGluGlnAsnVal 369
1158 TTCTTTATCGATTACGGCAAGCGCAACTCATCTTATCAACAAACATCA 1207
      |||||
369 ThrPheGluAsp.....AsnGlyThrLeuValLeuAsnGlnAsnIleA 384
1208 ACCAAGCGCGGGCGGCTTGTATTTGAAGTGATTTTACGGTCTCGCT 1257
      |||||
384 snGlnGlyAlaGlyLeuPhePheGlyAspTyrThrValLysGly 400
1258 GAAACAAACGAA...ACGTGGCAAGCGCGGCGCTTCATATCAGTGAAGA 1304
      |||||
401 AlaAsnAsnAspIleThrThrPheGlnGlyAlaGlyIleAspValAlaAspG1 417
1305 CAGTACCGCTTACTGGAAGTAAAGCGGTGGCAACGACGCGCTGTCCA 1354
      |||||
417 TyrLysValValAlaThrPheGlnValLysAsnProAsnGlyAspArgLeuAlaL 434
1355 AATCGGCAAGGCAAGCGCTGACGTTCAAGCCAAAGGGGAAACCAAGGC 1404
      |||||
434 ysIleGlyGlyThrLeuGlnIleAsnGlyThrGlyValAsnGlnGly 450
1405 TCGATCAGCGGGGCGGCGGTACGCTACGCTTGTGATCAGCAGCGAGCA 1454
      |||||
451 GlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnLysAlaAspAl 467
1455 TAAAGGCAAAACAAAGCCTTGTAGTAATCGCTTGTGNTCAGCGGCAAG 1504
      |||||
467 AspLysLysValGlnAlaPheSerGlnValGlyLeuValSerGlyArg 484
1505 GTACGGTGAAGTGAATGCGCATATCAGTTCACCCGACAACTGAT 1554
      |||||
484 LysThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuTyr 500
1555 TTCGGCTTTCGGGCGGCGGCTTGTGATTTAAAGGGCATTCGCTTCCGT 1604
      |||||
501 PheGlyPheAlaGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrPh 517
1605 CCACCGTATTCAAATACCGATGAAGGCGCGATGATTCNCATCAATATG 1654
      |||||
517 eGlnHisIleArgAsnValAspGlnGlyAlaArgIleValAsnHisAsnT 534
1655 CCACAACAATCCACCGCTTACCATTCACAGGAATGAAAGTATTACACAA 1704
      |||||
534 HisPheHisAlaSerThrIleThrLeuThrGlyLysSerLeuIleThrAsn 550
1705 CCGAGT..... 1710
551 ProAsnSerLeuSerValHisSerIleGlnAsnAspTyrAspGluAspAs 567
1711 .....GTAAGATA 1720
567 PylSerTyrTyrTyrArgProArgArgProIleProGlnGlyLysAspL 584
1721 TC.....AATGACTT 1731
      |||||
584 euTyrTyrLysAsnTyrArgTyrTyrAlaLeuLysSerGlyGlyArgLeu 600
1732 AAT..... 1734
      |||||
601 AsnAlaProMetProGluAsnGlyValAlaGluAsnAsnAspTyrPhePh 617
1735 .....TACAGCAAGAATAATGGC..... 1752
      |||||
617 emetGlyTyrThrGlnGluGlnAlaArgLysAsnAlaMetAsnHisLysAs 634
1753 .....TACAAAGTGGTGGCGAAGAAATACG 1782
      |||||
634 snAsnArgArgIleGlyAspPheGlyGlyPhePheAspGlnGluAsnGly 650
1783 ACCAAAGCAAGCGGCGCTCAACCTTGTATTCAGCGCCCGCAGAGA 1832
      |||||
651 LysGlnHisAsnGlyAlaLeuAsnLeuAsnPheAsnGlyLysSerAlaG1 667
1833 CCGCACCCNGTGCTTCCCGGCAACAATTTAAAGGCAACATCAGC 1882
      |||||
667 nLysArgPheLeuLeuThrGlyGlyAlaAsnLeuAsnGlyLysIleSerV 684
1883 AAACAAAGCGCAACTGTTTTCAGCGGACAGCGACCGACCGCCTTAC 1932
      |||||
684 alThrGlnGlyAsnValLeuLeuSerIleArgProThrProHisAlaArg 700
1933 AATCATTTAGAGAGCGGTGGTCAAAAATGGAAGT.....ATCCACA 1976
      |||||
701 AspPheValAsnLysSerSerAlaArgLysAspAlaHisPheSerLysAs 717
1977 AGGAGAAATGCTGGGACAAAGCAGCTGATCNCACCGCATTTAAACGG 2026
      |||||
717 nasGlnValValPheGluAspAspTyrPheAsnArgThrPheLysAlaA 734
2027 AAAATTCATATT...CAGGCGGCGGCGGCGGTGATTTCCCGCATGTT 2073
      |||||
734 laGluIleAlaValAsnGlnSerAlaSerPheSerSerGlyArgAsnVal 750
2074 GCCAAAGTGAAGCGCGATTGNCATTTGAGCAATCAGCCCAAGCAGTTT 2123
      |||||
751 SerAspIleThrAlaAsnIleThrAlaThrAspAsnAlaLysValAsnLe 767
2124 TGGTGTCGACCGCGCATCAAGCCATACATCTGTACAGCTTGGCGATGSA 2173
      |||||
767 uGlyTyr....LysAsnGlyAspGluValCysValArgSerAspTyr 782
2174 CNGCTGTGACAAATGTGTGCAANMAAACATTAACGACGTAAGTATT 2223
      |||||
782 hrGlyTyrValThrCys..... 787
2224 GCTTCATTGACTAAGACGACGACNCTNAGCGGACGANTGAGCTNNCCNATNA 2273
      |||||
787 ..... 787
2274 CGNTNTNTNMAANCTCNCNGGCGNTGCNCCATNAAGCGCATCTAGTG 2323
      |||||
788 .....AsnThrGlyAsnLeuSer. 793
2324 CAAATGGCATACACGTTATACAGTCAGCCGACCAAGCCACCAA...AAC 2370
      |||||
794 .....AspLysAlaLeuAsnSerPheAspAlaThrArgIleAsn 806
2371 GGCACCTTGAAGCTCGTGGGCAATGCCCAAGCAACATTTATATAGCCAC 2420
      |||||
807 GlnAsnValAsnLeuAsnGlnAsnAlaLeuValLeuGlyLysAlaAl 823
2421 ATTAAACGGCAACNATCGGNTTCGGGCAATGCTTATTAATCTAAGCA 2470
      |||||
823 aleuThrPheGlyLysIleGlnGlyGlnGlyAsnSerArgValSerLeuAsnG 840
2471 ACAAGCGCGCAAAAACGAGTGTGACGCTTTCGACCAACGCTAAGCA 2520
      |||||
840 LnhIs..... 841
2521 AACGTAAAGCATTCGCGACTCAAGCGCAATGTCTCCCTAGCGATAGCG 2570
      |||||
841 ..... 841
2571 AGTATTCATTTGAAACAGCGCGTTTACCGGCAACATCAGCGGCGAGCA 2620
      |||||
841 ..... 841
```

```

2621 AGGANAAGCATTTACACTTAAAGACAGCATGAGCGCTGCCCTCAGGC 2670
      |||:||||| ||| :||
842 ..... SerLysTrpHisLeuThrClyAsp 849
2671 ACGGAATTAGCAATTTAAACCTTGACAGCGCCACCATTTACATGATTC 2720
      |||:||||| ||| :|||
850 SerGlnValHisAsnLeuSerLeuAlaAspSerHisIleHisLeuAsnAs 866
2721 GCGCATGCGCCACGATGCGGAGCGGCAAAACCGCAGNGTGTCAGACA 2770
      |||:||||| ||| :|||
866 mAla..... SerAspAlaGlnSerAla..... 873
2771 CCGCGCGCGCGCTTCCGCGCGCTTCCATTTATCCGTTACACCGCAACT 2820
873 ..... 873
2821 TCGGTAGAAATCCCGTTTACACGCTGACGGGTAACGGCAATTGACNG 2870
      |||:||||| ||| :|||
874 ..... AsnLysTrpHisThrIleLysIleAsn...HisLeuSerG1 886
2871 TCAAGGAACATTCGCGTTATGTGCGAACTCTTCCGCTACGAGCGACA 2920
      |||:||||| ||| :|||
886 yAsnGlnHisPheHisTrpLeuThrAspLeuAlaLysAsnLeuGlyAspL 903
2921 AATTGAACTGGCGGAAAGTTCCAGAGNACTTACACCTTGCGCGCTAAC 2970
      |||:||||| ||| :|||
903 yValLeuValLysGlnSerAlaSerGlyHisTrpGlnLeuHisValGln 919
2971 AATACCGGCAAGAACCC.....GTAACTTCGATCAATTGACGGTGTGT 3014
      |||:||||| ||| :|||
920 AsnLysTrpGlyLuproAsnGlnLugLysAspLeuPheAspAlaSe 936
3015 GGAAGGAGAGACACAAACCCGCTCCGAAACCTTAATTTCACCCGCG 3064
      |||:||||| ||| :|||
936 rSerValGlnAspArg.....SerArgLeuPheValSerLeuA 949
3065 AAACGAGACAGCTGCATGCGCGCGCTTACCAACTCATCCGCAAA 3114
      |||:||||| ||| :|||
949 lAsnHisTrpValAspLeuGlyAlaLeuArgTrpThrIleLysThrGln 965
3115 GACGGCGAGTTCCGCTGCATTAAT.....CCGGT 3143
      |||:||||| ||| :|||
966 AsnGlyIleThrArgLeuTrpAsnProGlyAlaGlyAsnGlyArgProVa 982
3144 CAAGAGCAAGAGCTTCCGAAACCTCCGCAAGCGCAAGCGCAAAAAAC 3193
      |||:||||| ||| :|||
982 LysProAlaProSerProAlaAlaAsnThrAlaSerGlnAlaGlnLysA 999
3194 AGCGGGAAGAGACAGCG..... 3213
      |||:||||| ||| :|||
999 lAtThrGlnThrAspGlyAlaGlnIleAlaLysProGlnAsnIleValAl 1015
3214 ..... CAAGCCTTGACGCGCTGATTCGGCG 3239
1016 AlAProTProSerProGlnAlaAsnGlnAlaGlnLugLylAlaLeuArgGlnG1 1032
3240 CCGGCGCGATTCGCGCGGAAAGACAGAAAGCTTCCGCAACCGCGCGCG 3289
      |||:||||| ||| :|||
1032 nAlaLysAlaGlnLugLynValLysArgGlnIleAlaGlnAlaGlnLysV 1049
3290 NCGCAGCGCGGGAATGTGCGCATTTATGCGAGCGCGAGAGAGAAAAAA 3339
      |||:||||| ||| :|||
1049 aAla.....ArgGlnLysAspLugLylAlaLys 1058
3340 CCGGTTCAGCGCGAT..... 3354
      |||:||||| ||| :|||
1059 ArgLysAlaAlaGlnLylAlaArgGlnGlnLugLylAlaArgLysAlaAl 1075
3355 .....AAGACACAGCGNT 3367
      |||:||||| ||| :|||
1075 aGluLeuAlaAlaLysGlnLysAlaGlnAlaGlnLylArgLysAlaArgGlnL 1092
3368 TGGCGAACAAGCGCGAAGCGAACCAGCGCGGNTACACGCGCTTCCCG 3417

```

```

|||||:||||| |||:| :|||:| |||
1092 euAlaArgGlnLysAlaGlnLylAlaSerHisGlnAlaAsnAlaLysPro 1108
3418 CCGCGCGCGCGCGCGCGCGCGGATTTGCCGCAACG..... 3453
      |||:||||| ||| :|||
1109 LysArgArgArgAlaGlnAlaIleLeuProArgProProAlaProValPh 1125
3453 ..... 3453
1125 eSerLeuAspArgLysAlaLysAsnSerGlnUserSerIleGlyA 1142
3453 ..... 3453
1142 snLeuAlaArgValIleProArgMetGlyArgGlnLeuIleAsnAspTrp 1158
3453 ..... 3453
1159 GlnGlnIleProLeuGlnLugLysAspGlnAlaGlnLugLylArgAr 1175
3453 ..... 3453
1175 gGlnAlaThrGlnPheHisSerLysSerArgAsnArgArgAlaIleSers 1192
3453 ..... 3453
1192 eArgLuproSerAspGluAspAlaSerGlnSerValSerThrSerAsp 1208
3453 ..... 3453
1209 LysHisProGlnAspAsnTrpGlnLeuHisGlnLysValGlnThrAlaG1 1225
3454 .....CAAGCC.....CAAGCGAACCTCAACCCCAACCGCAGCGG 3490
      |||:||||| ||| :|||
1225 yLeuGlnProArgAlaAlaGlnProArgTrpGlnAlaAlaAlaGlnAla 1242
3491 ACCTGATNACCGCTTATGCCATAGCGGTTGATGTAATTTCCGCGACG 3540
      |||:||||| ||| :|||
1242 spAlaValSerThrAsnThrAsnSerAlaLeuSerAspAlaMetAlaSer 1258
3541 CTCACAGCGTT.....TTCGCGGTACGAGCAGCAATT 3572
      |||:||||| ||| :|||
1259 ThrGlnSerIleLeuLeuAspTrpGlnAlaTyrLeuThrArgHisIleAl 1275
3573 GAGCGCGCTGTTCCGAGAGCGCGCAACGCGTTTGACAAACGCGCA 3622
      |||:||||| ||| :|||
1275 aGlnLysSerArgAlaAspAlaGlnLysAsnSerValTrpMetSerAsnT 1292
3623 TCCGAGNACCAACAACTACCGCTTCCGAAATTTCCGCGCGCTACCGCGCA 3672
      |||:||||| ||| :|||
1292 hArgLyrGlnArgAspTrpAlaSerAlaGlnTrpArgArgPheSerSer 1308
3673 CAACGCACTGCGCGCAATCGGATGCAAGAAACCTCGGAGCGG... 3720
      |||:||||| ||| :|||
1309 LysArgTrpGlnTrpGlnIleGlyIleAspArgSerLeuSerGlnLysIme 1325
3721 CCGCGTGGCAATCCTGTTTCCGCAACCGGAGCGCAAAACANCTTGACAG 3769
      |||:||||| ||| :|||
1325 tGlnIleGlyGlyAlaLeuThrTrpTyrSerArgSerGlnHisThrPheAsp 1342
3770 ACGGCATCGGCACTCGGAGCGGCTTCCGAGCGCGCGCTTTCGCGCA 3819
      |||:||||| ||| :|||
1342 lAla...GlyGlyLysAsnThrPheValGlnAlaAsnLeuTyrGlyLys 1357
3820 TACGATCGCGG...AGTTGCAATCGGATGACAGCAGCGCGGCTTT 3866
      |||:||||| ||| :|||
1358 TyrTrpLeuAsnAspAlaTrpTrpValAlaGlyAspTrpIleGlyAlaGlySe 1374
3867 TACGAGCGGCAATCTNTCAGAGCGCATCGAGGCAAAATTCGCGCGCGG 3916
      |||:||||| ||| :|||
1374 rLeuArgSerArgLeuGlnThrGlnGlnLysAlaAsnPheAsnArgTrpS 1391
3917 TGTCTGCAATTACGCGATCAGGACAGATACCGCGCGGCTTTCGCGGATTC 3966
      |||:||||| ||| :|||

```

```

1391 eRLleGlnThrIleuThrlleuGlyAsnThrlleuLysIleasnGlnphe 1407
3967 GGCATGCAACCGTACATCGCGCAACGGCGCTATTGTCACAAAGGCGA 4016
1408 GlnIleValProSerAlaGlyIleArgTyrSerArgLeuSerSerAlaAs 1424
4017 TTACCGCTACGAAACGTCATATCGCCACCCCGGCTTGGCTTCACG 4066
1424 pYrIlySleuGlyAspSerValIlySerSerMetAlaValLysT 1441
4067 GNTACCGNCGGGCATTAAGCAGATATTTCATCAACCGCGCAACAC 4116
1441 hrlleuThrlleuGlyAspSerAlaTyrArgPheLys...ValGlyAsn 1456
4117 ATNTCCATCAGCCTTATTNAGCCTGTCCTATACCGATCGCGCTCGG 4166
1457 leuThrlleuThrlleuSerAlaAlaTyrPhe...AlaAsnTyrG 1472
4167 CAAGTCCGACACGCGTCATATCGCGCTATGCTCGATTCGCGCA 4216
1472 ylsGlyGlyValAsnValGlyIlySerPheAlaTyrLysAlaAsp 1489
4217 AACCCGCGATGCGGATGGCGGCTAACCGCGCAATCAAGCTTCACG 4266
1489 snGlnGlnIleThrlleuSerAlaGlyValAlaLeuLeuTyrArgAsnValThr 1505
4267 CTGTCCNTCCAGCTGCGCGCGCCCAAGGCGCAACTGAGGCGCAACA 4316
1506 leuAsnValAlaSnGlySerIleThrlleuGlyGlnLeuGlnLysGln 1522
4317 CAGCGCGGCGCATCAATTAGGCTACCGCTGCG 4347
1522 sserGlyGlnIleLysIleGlnIleArgPhe 1532

```

seq_name: SwissProt_40:IGA4_HAEIN

```

seq_documentation_block:
ID IGA4_HAEIN STANDARD; PRT; 1849 AA.
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NHRT HK61;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases."
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR: CLEAVES HOST IMMUNOGLOBULIN A
PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin a molecules at
certain pro-1-xaa bonds in the hinge region. No small molecule
substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6 (SERINE PROTEASES).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M67491; AAA24968.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolyase; Serine protease; Transmembrane; zymogen; signal.
FT SIGNAL 1 25
FT CHAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1022 1849 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 299 299 PROBABLE.
SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;

```

alignment_scores:

Quality: 1475.00 Length: 1924
Ratio: 1.517 Gaps: 64
Percent Similarity: 50.520 Percent Identity: 25.364

alignment_block:

US-09-303-518D-651 x IGA4_HAEIN ..

Align seg 1/1 to: IGA4_HAEIN from: 1 to: 1849

```

64 CGCTTCGCGCTGCTTACTATGACATATGCGCTGCTGCGCATCTTCG 113
5 LysPheLysLeuAsnPheIleAlaLeuThrlleuThrlleuThrlleuThr 21
114 CCAAGCTTGGGCGGACACACTTATTTCGCGCATCACTACCAATATCATC 163
21 oTyrThrlleuAlaAlaLeuValAlaArgAspAspValAspTyrGlnIlePhe 38
164 GCGACTTTCGCGCAAAATTAAGGCAAGTTTCAGTGGGCGGAAAGATATT 213
38 rGAspPheAlaGlnAlaSnLysGlyLysPheSerValGlyAlaThrAsnVal 54
214 GAGGTNTACAAACAAAAGGGAGTGTGGCGCAATCAATGACAAAGC 263
55 GluValAlaArgAspLysLysAsnGlnSerLeuGlySerAlaLeuProAsnG 71
264 C...CGATGATGATTTTCTGTGTGTG...CGTAACGGCGTGGCG 307
71 yIleProMetIleAspPheSerValAlaAspValAspLysArgIleAlaT 88
308 CATTTGTTGGCGCATCAATATATTTGTGAGCGTGACAT...AACGGCGC 354
88 hrlleuValAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
355 TATTAACAAGCTTGATTTTGGT...GCGGAGGAGAGNAA 389
105 ValSerGlnLeuIleSnPheGlyAsnLeuAsnGlyAsnMetAsnGlnGly 121
390 TCCGATCAGCAGCGT...TTTCTTACCAAAATG 421
121 nAlaLysSerHisArgAspValSerSerGluGluAsnArgTyrThrY 138
422 TGAAGAATAATATTAAGCTGACAT... 450
138 alGluLysAsnAsnPheProThrGlnAsnValThrSerPheThrThrLys 154
451 .....TCACACCTTACACAGCGGATTAACATATGCGCGTTT 488
155 GluGluGlnAspAlaGlnLysArgArgGluAspTyrTyrMetProArg 171
489 GCATAAATTTGTACAGATGAGAAAGCTGTGCA...ATGACGAGTGACA 535
171 uAspLysPheValThrGluValAlaProIleGluAlaSerThrAlaAsn 188
536 TGAGGGGGAATACCTATTCGATTAAGAAAATATCCGAGCGTGTCCG 585

```

```
188 snasnlysglyglutyrasnasnseraspstyrtproalaphavalarg 204
586 ATGGGCTGAGACACCATAT.....TGGGGTAT..... 615
205 leuglyserglyserlphenlelytyrlysglyserlthrglycine 221
616 .....GATGATGACAAACAGCGGATTTATCTACCGGCGAT 655
221 uilleuthrglylaspplsglnclyasnleuaurgasn.....T 236
656 GGTATTTGGCGGCATATACATATGACAGGTGGGAAATATATGCGCA 705
236 rpsaspvalglylasp...Asnleugluleuvalglyasnalarthyrthr 251
706 NTAGATTGAGCGGC.....GATGGCGGCATGCCAGCAC..... 741
252 TyrcllylelacllythrProtyrlysnvalasnhsiclyasnhsicle 268
742 .....TATGCG..... 747
268 uilleglypheglyasnserlysglunhsiseraspProlysglylel 285
748 .....CCTATGCCGATGACAGGTGGCGGACGACGCGTTCG 786
285 euserglaspProleuthrasnlyralavalleuglyaspserglyser 301
787 CCAATGTTTATGACAAACAAACAATAATGCTGCTACGACGAGT 836
302 Proleuphevalytrasparglyuglyllystrpleupheugly.. 317
837 TTTACAAACCGGCTACCTTATTCGGGAGGAAACGGTTTCACGCTGA 886
318 .....Sertyraspetherpalaclglyrasn..... 326
887 TAGCAAAAGTTGGTCTACGATGACATTTACAGAGCGCATACACTACC 936
327 ..lyslyssertrpInglytrpansilerylryshisglubhealgu 342
937 GTCNTTTTGAACCGGACGACGACATTTTCTTACATCCAAACA 986
343 lysiletyrglnglntyrserlalgly.....Serleuthrglyseras 357
987 C.....AACGGTACGGGTACGGTACAGAA 1012
357 nhrhrglntyrthrtrpGlnalarthrglyserthrserthrlethrglyg 374
1013 CCAACGAAAGTNTCCAAATCCAAAGCTTAAGTACAGACAGTCGACAG 1062
374 lylglylupProleu..... 378
1063 TTTGACGAATCTTTGAATGAATGAATGAAGACAGTTTACGCGGACAG 1112
379 .....SerValaspLeuthraspGly..... 385
1113 GGGTGTATACACTACCGTCCAAAGTTAAACAACGGTGAAACCTTTCTT 1162
386 .....LysaspLysProasnhsiclylysserlethr 397
1163 TTATGATTAACGCAACGCAACATCTTATCAAAACAACATCAACCA 1212
397 eu.....LysglySerglythrleuthrLeuasnhsisileaspGln 411
1213 GCGCGGCGGGTGTGATTTTGAAGTGAATTTTACGGTC...TCCGCTGA 1259
412 GlyAlaGlyGlyleuphehegluglyaspTyrglyvallysglythse 428
1260 AAACAACGAACGTCGCAAGCGCGGCGCTTCATATCACTGAAGACAGA 1309
428 rasperthrthrtrpLysglylalacllyalserValalaaspGlylyst 445
1310 CCGTTCCTTGGAAGTAAACGGCGGTGGCAACGACCGCTGTCCAAATC 1359
445 hvalthrtrpLysValHisasnProlystryparGlyleuAlaLysile 461

1360 GCAAGGACGACGTCGACGTTCAAGCCAAAGGGGAAACCAAGGCTGCAT 1409
462 GlylysglythrleuValaValaGlyuglyllysglyllysnnglyleule 478
1410 CAGCGTGGGCGAGGTACATGATTTTGGATTCACAGCAGCAGACGATAAG 1459
478 ulysvalaGlyaspGlythrValaIleleuLysGlnlyalalaspAlaasna 495
1460 GCAAAAAACAGCCTTAGTGAATTCGCTTGNTCAGCAGGAGGCTAGC 1509
495 snlysnvalaGlnalapheserGlnvalaGlylevalserGlylaryserthr 511
1510 GTGCACACTGATGCCATTAATCACTTCATCCACCAACACTCTATTCCG 1559
512 leuvalleuasnaspaspplsglvalalaspProasnserlethyrrpGcl 528
1560 CTTTCGCGCGGACGCTTGGATTTAAACGGGCAATTCGCTTCTTCACC 1609
528 yphetrglyllylaryleuaspplaspGlyAsnserleuthrphesph 545
1610 GTATTCAAAATACCGATGACGAGGCGGATGTCNMTCAATATGCCACA 1659
545 lsilearyasnileaspaspGlyalalargyvalaValasnhsisasmethr 561
1660 ACAACATCCACCGTTTACCATTAACAGGAAATGAATGATTAACACACCG.. 1707
562 AsnthrserasnlethrlethrglygluserleuthrlethrserProas 578
1707 ..... 1707
578 nthrlethrserlyrasnileGlnalaglnaspaspaspHisproleua 595
1707 ..... 1707
595 rgllearyserleProtyrargluleuthrphesnglnaspasnarg 611
1707 ..... 1707
612 serlyrlythrleuLyslyglYAlaserthrargserguleuprogl 628
1708 ...AGTGGTAAAGAT..... 1719
628 nasnserylgluserasnglnaSntrpleuthrmetGlylarythrsera 645
1720 .....ATCATGACTTATTTACAGAAAGAAAT 1749
645 spglualalysargasnValmetasnhsisileasnnglnalargmetasn 661
1750 GCCTACAAAGGTTGGTTGGGAGAAAGATPACGACCAAAACG...AACGG 1796
662 Glyphesnglyllyrphenglygluglngluthrlysnlathrglnasngl 678
1797 GCGGCTCACCTTGTTTACACGCCCGCGCAAGAGACCGACCCGCTGC 1846
678 llylsleuasnvalathrphesnglyllyseraspGlnasnhrpHeuleu 695
1847 TTCCGCGGGAACAATTTAAAGGCAACATCAGCAACAAACAGGCAAA 1896
695 euthrglyglythrAsnleuasnnglylaspLeuasnValaGlyllyglythr 711
1897 CTGTTTTTACGGGCGACGACGACGACCGGACGCTTACATTTAGAGAG 1946
712 leupheuserglylarygprothrProHisAlaargaspIlealaglyl 728
1947 CGGGTGTCAAAATGAAGGT.....ATCCCAAGAGGAAATCTGCT 1990
728 eserSerthrlyllysnaspProHisphethrGlnaasnnglnalvalaVal 745
1991 GGGCAACGACTGATCNACCGCAGCTTTAAAGCGGAAATTTCCATATT 2040
745 alGlnaspaspTrpIleasnargasnpheslysalatnhrmetasnval 761
```



```
2041 CAGGCGGGGACGCGGTGATTTCC...CGCAATGTTGCCAAAGTGGAAG 2087
||||| : : : : : ||||| : : : : :
762 ThrGlyAsnAlaSerLeuTyrSerGlyArgAsnValAlaAsnIleThrSe 778
||||| : : : : : ||||| : : : : :
2088 CGATTGNCATTGAGCATTCAGCCGACAGAGTCTTTGGTGTGCACCGC 2137
||||| : : : : : ||||| : : : : :
778 rAsnIleThrAlaSerAsnAsnAlaGlnValAlaHisIleGlyTyr...L 793
||||| : : : : : ||||| : : : : :
2138 ATCAAGACCATCATCTGTACAGCTGCGAGCTGGACNGGCTGCAANT 2187
||||| : : : : : ||||| : : : : :
793 ysthGlyAspThrValCysValArgSerAspTyrThrGlyTyrValThr 809
||||| : : : : : ||||| : : : : :
2188 TGTGTGGAANAACATTACCGAGCATTAAGTATGCTTCATGTACTAA 2237
||||| : : : : : ||||| : : : : :
810 CysHisAsnSerAsnLeuSer...GluLysAlaLeuAsnSerPheAsnPr 825
||||| : : : : : ||||| : : : : :
2238 GACNACACMTNAGCGCANTGTNAGCTNNCCNATNACNTNNTTAAANC 2287
||||| : : : : : ||||| : : : : :
825 oThrAsnLeuArgGlyAsnValAsnLeuThrGluAsnAlaSerPheThrL 842
||||| : : : : : ||||| : : : : :
2288 TCNCGGCGCNTGCTNACCTNNAANGCATCTTACGCAATGGCGATACA 2337
||||| : : : : : ||||| : : : : :
842 eu...GlyLysAlaAsnLeuPheGlyThrIleGlnSerIleGlyThrSer 857
||||| : : : : : ||||| : : : : :
2338 CGTTATACAGTCAGCCACACAGCC...ACCCAAACGGGCA 2375
||||| : : : : : ||||| : : : : :
858 GlnValAsnLeuLysGluAsnSerHisTrpHisLeuThrGlyAsnSerAs 874
||||| : : : : : ||||| : : : : :
2376 CCTTACGCTCGTGGC...AATGCCCA 2400
||||| : : : : : ||||| : : : : :
874 nValAsnGlnLeuAsnLeuThrAsnGlyHisIleHisLeuAsnAlaGln 891
||||| : : : : : ||||| : : : : :
2401 .....GCAACATTTATCAAGCCACATTAAGCGCAC 2433
||||| : : : : : ||||| : : : : :
891 snAspAlaAsnLysValThrThrTyrAsnThrLeuThrValAsn... 905
||||| : : : : : ||||| : : : : :
2434 NCATCGGNTCCGGCAATGCTTCATT... 2460
||||| : : : : : ||||| : : : : :
906 ...SerLeuSerGlyAsnGlySerPheTyrTyrTrpValAspPheThrAs 921
||||| : : : : : ||||| : : : : :
2461 .AATGTAGCAACAGCGCCGCAAAACGGCAGT... 2493
||||| : : : : : ||||| : : : : :
921 nasLysSerAsnLysValValValAsnLysSerAlaThrGlyAsnPheT 938
||||| : : : : : ||||| : : : : :
2494 .....CTGACG 2499
||||| : : : : : ||||| : : : : :
938 hLeuGlnValAlaAspLysThrGlyLupProAsnHisAsnGluLeuThr 954
||||| : : : : : ||||| : : : : :
2500 CTTTCGACAAAGCTAAGCA...AACGTAAAGCCATTCCGCACT 2540
||||| : : : : : ||||| : : : : :
955 LeuPheAspAlaSerAsnAlaThrArgAsnAsnLeuGlnValThrLeuAl 971
||||| : : : : : ||||| : : : : :
2541 CAGCGCATGTCTCCCTAGCCGATTAAGCAGTATTCATTTGAAAAACA 2590
||||| : : : : : ||||| : : : : :
971 asnGlySerVal...AspArgGlyAlaThrLysTyrLysLeuA 985
||||| : : : : : ||||| : : : : :
2591 GCCGCTTACCGACACTCAGCGCGCAGCAAGANACAGCATTTACACTTA 2640
||||| : : : : : ||||| : : : : :
985 rGAsnValAsnGlyArgTyrAsp...Leu 993
||||| : : : : : ||||| : : : : :
2641 AAAGACAGCAATGAGCGCTGCCGTACAGCAGCAATTAAGCAATTTAA 2690
||||| : : : : : ||||| : : : : :
994 TyrAsnProGlu...ValGluLysArgAsnGlnThr 1004
||||| : : : : : ||||| : : : : :
2691 COTTACAAACGCCACCATTAACATTCGCGCTATGCCACAGTGTG 2740
||||| : : : : : ||||| : : : : :
1004 rValAspThrThrAsnIleThrThrProAsnAspIleGlnAlaAspAlaP 1021
||||| : : : : : ||||| : : : : :
2741 CAGCGCGCAACAGCGCAGNGTGTCA...GACACGCGCGC 2778
||||| : : : : : ||||| : : : : :
1021 roSerAlaGlnSerAsnAsnGluGluIleAlaArgValGluThrPro... 1036
||||| : : : : : ||||| : : : : :
2779 CGCGCTTGGCGCGTCCCTATTATCGGTACACCGCCAACT... 2820
||||| : : : : : ||||| : : : : :
||||| : : : : : ||||| : : : : :
1037 .....ValProProAlaProAlaThr 1044
||||| : : : : : ||||| : : : : :
2821 .....TCGTTAGATCCCGTTTCAACACAGCTGACGGTAAAG 2857
||||| : : : : : ||||| : : : : :
1044 rGluSerAlaIleAlaSerGluGln...ProGluThrArgProAlaGluThr 1060
||||| : : : : : ||||| : : : : :
2858 GCAATGTGACNGTCAGAAACATTC...GCTT 2889
||||| : : : : : ||||| : : : : :
1061 AlaGlnProAlaMetGluGluThrAsnThrAlaAsnSerThrGluThrAl 1077
||||| : : : : : ||||| : : : : :
2890 ATGTGGAACCTTTCGGCTACCGAAGCGCAAAATTGAAGCTGGCGAAG 2939
||||| : : : : : ||||| : : : : :
1077 aProLysSerAspThrAlaThrGlnThrGluAsnProAsnSerGluSerV 1094
||||| : : : : : ||||| : : : : :
2940 TTCGGAAGNACTTACA...CCTTGGCGGTCA 2968
||||| : : : : : ||||| : : : : :
1094 aLProSerGluThrThrGluLysValAlaGluAsnProProGlnGluAsn 1110
||||| : : : : : ||||| : : : : :
2969 ACAATTACCGCAACGCGCTAAGCCTGATCATTAATTGACGTAGTGAA 3018
||||| : : : : : ||||| : : : : :
1111 GluThrValAlaLysAsnGluGlnGluAlaThrGluProThrProGlnAs 1127
||||| : : : : : ||||| : : : : :
3019 GGGAAAGACAACAACCGCTGTCCGAAACCTTAATTTCACCCGCAAA 3068
||||| : : : : : ||||| : : : : :
1127 nGlyGluValAlaLysGluAsnProThrValGluAlaAsnThrGlnT 1144
||||| : : : : : ||||| : : : : :
3069 CGAACACGTGATGCCGCGCGGTGCTAC... 3100
||||| : : : : : ||||| : : : : :
1144 hAsnGluAlaThrGlnSerGluGlyThrGluGluThrGlnThrAla 1160
||||| : : : : : ||||| : : : : :
3101 .....AACATC 3108
||||| : : : : : ||||| : : : : :
1161 GluThrLysSerGluProThrGlnSerValThrValSerGluAsnGlnPr 1177
||||| : : : : : ||||| : : : : :
3109 CGCAAGACGCGAGTTCGCGC...TCGATATCCGCTCA 3146
||||| : : : : : ||||| : : : : :
1177 oGluLysThrValSerGlnSerThrGluAspLysValValGluLysG 1194
||||| : : : : : ||||| : : : : :
3147 AGAACAAAGAC...TTCCGACAAAC 3169
||||| : : : : : ||||| : : : : :
1194 LuGluLysAlaLysValGluThrGluGluThrGlnLysAlaProGlnVal 1210
||||| : : : : : ||||| : : : : :
3170 TCGCAAGCGCAGAAAGCCAAACAGCGCA... 3201
||||| : : : : : ||||| : : : : :
1211 ThrSerLysLupProPro...LysGlnAlaGluProAlaProGluGluVal 1226
||||| : : : : : ||||| : : : : :
3202 .....AAAGACAAGCGCAAGCCTTGACGCGCTG... 3231
||||| : : : : : ||||| : : : : :
1227 ProThrAspThrAsnAlaGluGluAlaGlnAlaLeuGlnGlnThrGlnPr 1243
||||| : : : : : ||||| : : : : :
3232 .....ATTGCGCGCGCGCGCATGCC...GCCG 3256
||||| : : : : : ||||| : : : : :
1243 oThrThrValAlaAlaAlaGluThrThrSerProAsnSerLysProAlaG 1260
||||| : : : : : ||||| : : : : :
3257 AAAAGACAGAA...AGCGTCCGACCGCGCGCGCNG 3291
||||| : : : : : ||||| : : : : :
1260 LuGluThrGlnGlnProSerGluLysThrAsnAlaGluProValIlnPro 1276
||||| : : : : : ||||| : : : : :
3292 GCAAGCGGGGAAAAATGTCCGATTAATGACAGCGGAG...GAAAGAAAA 3338
||||| : : : : : ||||| : : : : :
1277 ValValSerGluAsnThrAlaThrGlnProThrGluThrGluGluThrAl 1293
||||| : : : : : ||||| : : : : :
3339 ACGGTGCGAGCGGATTAAGACAGCCNTTG... 3369
||||| : : : : : ||||| : : : : :
1293 aLysValGluLysGluLysThrGlnGluValIProGlnValAlaSerGlnG 1310
||||| : : : : : ||||| : : : : :
3370 .....GCGAAGACGCGGAGCGGAAACCGCGCGGNTACACCGCTTC 3414
||||| : : : : : ||||| : : : : :
1310 LuSerProLysGlnGluGlnProAlaAlaLysProGlnAlaGlnThrLys 1326
||||| : : : : : ||||| : : : : :
3415 CCCCGCGCGCGCGCGCGCGGATTTG... 3444
||||| : : : : : ||||| : : : : :
```

```

1327 ProGlnAlaGluProAlaArgGluAsnValLeuThrThrLysAsnValGln 1343
3445 ....CCGCAACCGCAGCCCAACCGCA..... 3468
||||| |||||
1343 yGluProGlnProGlnAlaGlnProGlnThrGlnSerThrAlaValProT 1360
3468 ..... 3468
1360 hrThrGluThrAlaAlaAsnSerLysProAlaAlaLysProGlnAla 1376
3469 .....CCGCAACCGCAGCCGAGCGGCTGATNAG..... 3501
||||| |||||
1377 GlnAlaLysProGlnThrGluProAlaArgGluAsnValSerThrValAs 1393
3501 ..... 3501
1393 nrThrLysGluProGlnSerGlnThrSerAlaThrValSerThrGluGlnP 1410
3501 ..... 3501
1410 roAlaLysGluThrSerSerAsnValGluGlnProAlaProGluAsnSer 1426
3501 ..... 3501
1427 IleAsnThrGlySerAlaThrThrMetThrGluThrAlaGluLysSerAs 1443
3501 ..... 3501
1443 pLysProGlnMetGluThrValThrGluAsnAspArgGlnProGluAlaA 1460
3501 ..... 3501
1460 snThrValAlaAspAsnSerValAlaAsnAsnSerGluSerSerGluSer 1476
3502 .....CGTATGCCAATACGGCTTGAGTGAATTTCCGCCAC 3539
||||| |||||
1477 LysSerArgArgArgSerValSerGlnProLysGluThrSerAlaGln 1493
3540 GCTCAACAGCGTTTCGCCGTACAGAC..... 3567
||||| |||||
1493 uGluThrThrValAlaSerThrGlnGluThrThrValaAspAsnSerValS 1510
3567 ..... 3567
1510 erThrProLysProArgSerArgThrArgArgSerValGlnThrAsn 1526
3568 .....GAATTGACCGCGCTTTGCCGAGACCGCGCA 3602
||||| |||||
1527 SerThrGluProValGlnLeuProThrGluAsnAlaGluAsnAlaGluAs 1543
3603 CGCN..... 3606
|||
1543 nValGlnSerGlyAsnAsnValAlaAsnSerGlnProAlaLeuArgAsnL 1560
3606 ..... 3606
1560 euThrSerLysAsnThrAsnAlaValLeuSerAsnAlaMetAlaLysAla 1576
3606 ..... 3606
1577 GlnPheValAlaLeuAsnValGlyLysAlaValSerGlnHisIleSerGln 1593
3607 .....GTTTGACACAGCNCATCC 3625
||||| |||
1593 nLeuGluMetAsnAsnGluGlyGlnIleThrAsnValIlePheSerAsnThrS 1610
3626 GGNACACCAACACTACCGTTGCAAGATTTCGCGCTACCGCAACAA 3675
||||| |||||
1610 erMetAsnLysAsnThrSerSerGluGlnIleThrArgArgPheSerSerLys 1626
3676 ACCGACCTCGGCAATCGTATGCAAGAAACCTCGGAGCGG...CG 3722
||||| |||||
1627 SerThrGlnThrGlnLeuGlyIleThrAspGlnThrIleSerAsnAsnValGln 1643

```

```

3723 CGTGGCATCTCTTTTCGCACACCGGACCGAAACACTTCGACGACG 3772
||||| |||||
1643 nLeuGlyGlyValPheThrThrValArgAsnSerAsnAsnPheAspLysA 1660
3773 GCATCGGCACACTCGGCACGCGCTTCCACGCGCGCTTTTCGGCAATAC 3822
||||| |||||
1660 lAsrSerLysAsnThrLeuAlaGlnValAlaAsnPheThrLysThrLys 1676
3823 GGCATCGGCAGCTTCGACATCGCATACGACGCGCGCGCTTTTACACG 3872
||||| |||||
1677 AlaAspAsnHisThrPheThrLysGlyIleAspLeuGlyThrLysPheGln 1693
3873 CGGCATCTTTCACAGCAGCGCATCGGAGCAAAATCGCGCGCGCTGTC 3922
||||| |||||
1693 nSerAsnLeuGlnThrAsnAsnAlaLysPheAlaArgHisThrAlaG 1710
3923 ATTACGGCATTCAGGACGACGATACCGCGCGCTTTCGGCGGATTCGCGATC 3972
||||| |||||
1710 nIleGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheAlaVal 1726
3973 GACCGTACATCGCGCGCAGCGCGCTATTTCGTCACAAAGCGGATACCG 4022
||||| |||||
1727 LysProThrValGlyValArgThrSerThrLysSerAsnAlaAspPheAl 1743
4023 CTACGAAACGTCATATCGCCACCGCGCTTCGCTGACCGCATACCG 4072
||||| |||||
1743 aLeuAlaGlnAspArgIleLysValAsnProIleSerValLysThrAlaP 1760
4073 GNGCGGCATTAAAGCAGATTATTCATTCACACCGCGGCAACACATTC 4122
||||| |||||
1760 nAlaGlnValaAspLeuSerThrThrThrHisLeuGlyGlu...PheSer 1775
4123 ATCAGCCTTATTNAGCCTCTCTATACCGAGCCGCTTCGGGCAAGT 4172
||||| |||||
1776 lIleThrProIleLeuSerAlaArgTyr...AspAlaAsnGlnGlyAsnGln 1791
4173 CCGAACAACGTCATATCGCCACCGCGCTTCGCTGACGATTCGCGCAAA 4222
||||| |||||
1791 yLysIleAsnValSerValIleThrAspPheAlaThrValAsnValGluAsnGln 1808
4223 GCAGTGGCGAATGGCGGCTTAACCGCGCAATCAAGGTTTCAGCGCTCC 4272
||||| |||||
1808 lGlnIleThrAsnAlaGlyLeuLysLeuLysThrHisAsnValLysLeuSer 1824
4273 NTCACGCTGCGCGCGCAAGCGCAAGCTGAGACGCAACACGACGCG 4322
||||| |||||
1825 LeuIleGlyGlyLeuThrLysAlaLysGlnAlaGlnLysGlnLysThrAl 1841
4323 GGCATCAATTAAGCTAC 4341
||||| |||||
1841 aGluValLysLeuSerPhe 1847
seq_name: SwissProt_40:YPIA_ECOLI
seq_documentation_block:
ID YPIA_ECOLI STANDARD; PRT: 1569 AA.
AC P52143; P76610; P77017; P77019;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical outer membrane protein yPjA.
GN YPIA OR B2647.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

```



```

480 TyrThrGlyGly.....IleAlaHisGlyValAlaSerGlnGluThrGln 493
1221 CCGTTGTATTTTGAAGGTGATTTTACGGTCTCGCTGAAACAAACGAAA 1270
      |||:::
493 ySerAlaLeuValAlaAsn..... 499
1271 CGTGGCAAGCGCGCGCTTCATATACGT.....GAAGAC 1305
      |||:::
500 .....ThrGlyAlaGlyThrAspIleGluGlyTyrAsnLysLeuSerHis 514
1306 AGTACGTTACTGTGAAAGTAACGGCGTGGCAAGCAGCCGCTGTCCAA 1355
      |||:::
515 PheThrIleThr.....GlyGlyGluAlaAsn..... 523
1356 AATGGCAAAAGCAGCGTCGACGCTTCACGCAAAAGGGAACCAAGCGCT 1405
      |||:::
524 .....TyrValValLeuGluAsnThrGlyG 532
1406 CGATCAGCGTGGCGAGCGGTACATCTTTGGATCAGCAGCAGACGAT 1455
      |||:::
532 IuLeuThrValValAlaLysThrSerAlaLysAsnThrThrIleAspThr 548
1456 AAGGCAAA.....AACAAGCCTTTAGTGAATCGGCTTGN 1493
      |||:::
549 GlyGlyLysLeuIleValGlnLysGluAlaLysThrAsp..... 561
1494 CAGCGCAGGGGCTACGCTGCAACTGAATCGCATATCACTCAACCCCG 1543
      |||:::
562 .....SerThrArgLeuAsn..... 566
1544 ACAACTGTATTTTGGCTTCGGCGGCGAGCGTTGGATTAAACGGCGAT 1593
      |||:::
567 .....AsnGlyGlyValLeuGluValAlaGlnAspGly 576
1594 TCGCTTTCGTCCACCGTATTCAAATACCAGTGAAGGGCGCATATGTCN 1643
      |||:::
577 GlyGluAlaLysHis...ValGlnGlnLysGlyAlaLeuIle... 591
1644 CNACTAATAGCCACAAACATCCACCGCTTACCATTCACGGAATGAAA 1693
      |||:::
592 .....AlaSerThrThrSerGlyThrLeuIleGlnGlyThrAsnS 605
1694 GT.....ATTACAAACGAGTGGTGAATATCAAT 1725
      |||:::
605 eTyrGlnLysAspAlaPheTyrIleArgAsnSerGluAlaLysAsnValAl 621
1726 AGACTTAATTTACAGCAAGAATTTGGCTACAGCGTGGTGGCGAGAA 1775
      |||:::
622 LeuGluAsnAlaGlySerLeuThrValValThrGlySerArgAlaValAs 638
1776 AGATACGACCAAAACGACGCGGCTCAACCTT..... 1809
      |||:::
638 pThrIleIleAsnAlaAsnGlyLysMetAspValTyrGlyLysAspValG 655
1810 .....GTTTACCAGCCGCCCGCA 1827
      |||:::
655 IYThrValLeuAsnSerAlaGlyThrGlnThrIleTyrAlaSerAlaThr 671
1828 GAAGACCGCAGCCNGCTGCTTCGGCGGCAACAATTTAAAGGACACAT 1877
      |||:::
672 SerAspLysAlaAsnIleLysGlyLysGlnThrValTyrGlyLeuAl 688
1878 CACGCAAAACAAAC.....GGCAAACTGTTTTCAGCGGCAAGCCGA 1918
      |||:::
688 aThrGluAlaAsnIleGlnSerGlyGlnIleValAlaSpGlyLysert 705
1919 CACGCAAGCCCTACATCATTTAGAGAGCGGCTGCTCAAAAATGGAAGT 1968
      |||:::
705 hGlnLys.....ThrHisIleAsnGlyGlyThrGlnThrValGlnAsn 719
1969 ATCCCAACAAGA.....GAATCGTGTGGAGCAACGACTGATGATCA 2009
      |||:::
720 TyrGlyLysAlaIleAsnThrAspIleValSerGlyLeuGlnIleLeu 736
2010 CCGCAGTTTAAACGGGAAATTTCCATATTCAGGCGCGGCGGCGTGA 2059
      |||:::
736 tAlaAsnGlyThrAlaGlnGlySerIleIleAsnGlyLysertInValY 753
2060 TTTCGCCG..... 2067
      |||:::
753 aAlaSerGlnGlyGlyLeuAlaGlnAsnSerValLeuAsnAspGlyGlyThr 769
2068 .....AATGTGCCAAGTGGAGGCGGATTTGNCATTTGAGCAATCGGCCA 2114
      |||:::
770 LeuAspValArgGlnLysGlySerAlaThrIleGlnGlnSerGln 786
2115 AGCAGTTTTCGTCCGACCGCATCAAGCAATCATCTGTACAGCTT 2164
      |||:::
786 nGlyAlaLeuValAlaThrThrArgAlaThrArgValThrGlyThrArg 803
2165 CGAGCTGGACN.....GCTGTGCAAAAT..... 2187
      |||:::
803 IAspGlyValAlaPheSerIleGlnGlnGlyAlaAlaAsnIleLeu 819
2188 .....TGTGTGAAANCAATTCACGCGATTA 2216
      |||:::
820 LeuAlaAsnGlyGlyValLeuThrValGlnLysSerAspThrSerAsp 836
2217 AGTATTGCTTCATTTG.....ACTAAGCAGCAGCN 2245
      |||:::
836 sThrGlnValAsnMetGlyArgGlnIleValLysThrLysAlaThr 853
2246 TNAGCGGCANTGTNAGCNCTNACNATNAGTNTTAAANCTCNCNGG 2295
      |||:::
853 IatnGly.....ThrThrLeuThrGlyGly 861
2296 CNTGCNACCTNANAGC..... 2313
      |||:::
862 GluGlnIleValGlnGlyValAlaAsnGlnThrThrIleAsnAspGly 878
2314 .....AATCTAGTCAAAATGGCGAT.....ACAGTTATACAG 2347
      |||:::
878 yIleGlnThrValSerAlaAsnGlyGluAlaIleLysThrLysIleAsn 895
2348 TCAGCCCAACGCCCAACGCGCAAC.....CTTACGCTCGTG 2388
      |||:::
895 IuGlyGlyThrLeuThrValAlaAsnAspAsnGlyLysAlaThrAspIleVal 911
2389 GGCATGCCCCAGACATTTAATCAGCCCAATTAAGGCAACNATNC 2438
      |||:::
912 GlnAsnSerGlyAlaAlaLeuGlnThrSerThrAlaAsnGly...IleG 927
2439 GGNFTTCGGGCAT.....GCTTCATTTAATCTAAGCAACAGCCG 2479
      |||:::
927 uIleSerGlyThrHisGlnIleGlyThrPheSerIleSerGlyAsnLeu 944
2480 CACAAC.....GGCAGTGTGACGCTTTCGCAAC 2511
      |||:::
944 IatnAsnMetLeuLeuGlnAsnGlyGlnAsnLeuValLeuAlaGly 960
2512 GCTAAGCAAAAGTAAAGCAATTCGCGACTCAAGGCAATTCCTCCATAG 2561
      |||:::
961 ThrGluAlaArgAspSerThrValGlyLysGlyGly... 972
2562 CGATNAGCAGTATTCATTTTGAACACAGCCCTTACCGCAACTCA 2611
      |||:::
973 .....AlaMetGlnAsnLeuGlnAspSerAlaThrLysValAsn 987
2612 GCGGCAAGCAAGANACAGCATTTACACTTAAAGACAGCAATGAGCGTG 2661
      |||:::
987 eGlyGlyGlnIleYThrLeuGlyArgSerLysAspGlnPheGlnAlaLeu 1003
2662 CCGTCAAGCAGCAATTA..... 2679
      |||:::
1004 AlaArgAlaGlnAspLeuGlnValAlaGlyGlyThrAlaIleValTyrAl 1020

```

2680	GGCAATTAAAC	2692
1020	aglythrleuAlaaspAlaSerValSerGlyAlaThrGlySerLeuSerL	:	1037
2693	TT.....		2694
1037	eueThrProAtrgAspAsnValThrProValLysLeuGluGlyAlaVal		1053
2695	GACACGCCACCAATTACACTCAATTCGCCCTAT.....	2727
1054	ArgIleThrAspSerAlaThrLeuThrLeuGlyAsnGlyValAspThrTh	:	1070
2728	.CGCCACGATGTCGCGGGGGCCAAACCGGCGAGNTG.....	:	2763
1070	rleuAlaaspLeuThrAlaAlaSerTrgGlySerValThrIleuAsnSera		1087
2764	TCAGACGCGCGCGCGCGCTTCGCG...CGTCCCTA	2799
1087	snsnSerGysAlaGlyThrSerAsnGysGlyThrArgValAsnSerLeu	:	1103
2800	TTA.....		2819
1104	leuIleuAsnAspGlyAspValTyrlleuSerAlaGlnThrAlaAlaProAl	:	1120
2820	TTGGGTAGAAATCCCGTTACAACGTCGAGTAAACGGCAAAATTGAA		2866
1120	atrrThrAsnGlyIleThrAsnThrleuThrAsn...GluLeuSerG	:	1136
2870	GTCAGAGAACATTCGCCCTTATGTCGGAACCTTGGGTACCGAAGC		2919
1136	lyserGlyAsnPhetYrleuHisThrAsnValAlaGlySerTrgGlyasp	:	1155
2920	AAATTGAAGCTGGCGGAAGATTCCGAAGGAACTACACCTGGCGGTCA		2965
1153	GlyleuValValAsnAsnAsnAlaThrGlyAsnPhetYrIlePheValG	:	1169
2970	CATPACCGGCAACGACCGGTAAACCTGATCAATTGACGGTAGAGAG		3019
1169	nasPThrGlyValSerProGlnSerAspAspAlaMetThrleuVal...	:	1184
3020	GGAAGAACAACAACCGCTGTCCGAAACCTTAATTTCACCTCGCAAA		3063
1185	:	1197
3070	GAA.....	CAGTCGATGCGCGCGCGGTTCACCACTCATCCGCA	3113
1198	ThrGlyGlyPheValAspLeuGlyThrTyrlGlyTyrl...ValleuLys	:	1213
3114	AGACGGC.....	GAOTTCGCGCTCATATACTCCGTCGCAAGAACAGAC	3157
1213	rAspGlyIleAsnSerAsnTrpAsnleuThrAsnAspAllys.....	:	1226
3158	TTTCCGACAAACTCGGCAAGGCAAGCCAAAAAAACAGCGGCAAAAA		3207
1226		1226
3208	AACGCGCAAGCCTTGACGCGCTGATTGCGCGGGCGGATGCCCGCA		3257
1227	Pr	1227
3258	AAAGACAGAAAGCGTTGGCCGAACCGCGCGGCGAGCGGGGAAATG		3307
1227	oAsnProAspProIleProAsnPro.....		1235
3308	TCGGCATTTATGCAAGGGGAGAGAAAGAAAAACGGTGCAGCGGATAA		3357
1235		1235
3358	GACAGCGCNTTGGCGAAGACGCGCAAGCGCAACCGCGCGGNTACCAC		3407
1236	LysProAspProIlyAspProAspProly	1244
3408	CGCGTTCCCGCGCGCGCGCGCGCGGATTTGGCCGCAACCGCGAC		3457

[illegible]

```

1515 pGlyAsnPhenValAsnTyrLeuSerGlyArgArgGlyIleTyrGlnAlaG 1532
4238 GCGTAAACGCCGAATCAAGGTTTCACGCTGTCACGTCACGCGCCG 4287
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1532 LyleIySalaserPheSer...ThleuSerGlyHisLeuGlyVal 1547
4288 GCCAAAGCCGCACTGCAAGGCAACAGCGCGGATCAAA..... 4332
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1548 Gly.....TyrSerHisSerAlaGlyValGluSerPr 1558
4333 .....TTAGCTACCGCTGG 4347
1558 OTRPAsnAlaValAlaGlyValAsnTyr 1567

seq_name: SwissProt_40:BIGA_SALTY

seq_documentation_block:
ID BIGA_SALTY STANDARD; PRT; 1953 AA.
AC P25927; P25928; Q9XCQ3;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Putative surface-exposed virulence protein biga precursor.
GN BIGA OR STM3478.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=14028;
RA Stojiljkovic I., Valentine P., Heffron F.;
RT "Salmonella typhimurium rhs homolog.";
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., All J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE OF 1-765 FROM N.A.
RC STRAIN=LT2;
RA MEDLINE=91100301; PubMed=1987123;
RA Wu J.Y., Siegel L.M., Kredich N.M.;
RT "High-level expression of Escherichia coli NADPH-sulfite reductase:
RT requirement for a cloned cysG plasmid to overcome limiting siteheme
RT cofactor";
RL J. Bacteriol. 173:325-333(1991).
RN [4]
RP CAUTION: Ref.3 sequence differs from that shown due to frameshifts
RP in positions 414 and 732.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL; AF133696; AAD39458.1; -
DR EMBL; AE008859; AAL22340.1; -
DR EMBL; M64606; AAA27042.1; ALT_FRAME.
DR EMBL; M64606; AAA27043.1; ALT_FRAME.
DR PIR; C39200; C39200.
DR PIR; D39200; D39200.
DR StyGene; SG10437; BIGA.

```

```

KW Virulence; Repeat; signal; Complete proteome.
FT SIGNAL 27
FT CHAIN 28 1953
FT DOMAIN 101 252
FT REPEAT 101 103
FT REPEAT 104 113
FT REPEAT 114 122
FT REPEAT 123 133
FT REPEAT 134 144
FT REPEAT 145 155
FT REPEAT 156 166
FT REPEAT 167 177
FT REPEAT 178 188
FT REPEAT 189 199
FT REPEAT 200 210
FT REPEAT 211 221
FT REPEAT 222 232
FT REPEAT 233 243
FT REPEAT 244 252
FT REPEAT 207 207
FT CONFLICT 514 514
FT CONFLICT 1698 1698
FT CONFLICT 1795 1798
FT CONFLICT 1836 1837
SQ SEQUENCE 1953 AA; 200150 MW; 611B3F1C954D91AE CRC64;

```

```

alignment_scores:
Quality: 264.50 Length: 1669
Ratio: 0.359 Gaps: 74
Percent Similarity: 44.158 Percent Identity: 18.394

```

alignment_block:

US-09-303-518d-651 x BIGA_SALTY ..

Align seg 1/1 to: BIGA_SALTY from: 1 to: 1953

```

217 GTNACACAAACAAAGGAGGAGTGTGCGCAATCAATGACAAAGCCCC 266
||| ||| ||| |||:|||||: ||| ||| |||
413 ValAspAsnAlaThrGlyAlaLeuIleSerGlyAsnGlyThrThrThr 429
267 GATGATTGATTTTCTGTGTCGCGCTAACGCGGCGCATGGTG. 315
::: ||||| |||||:|||||: ||| ||| |||
429 nPheAlaGlyAspIleAlaValSerGlyGlyThrAlaIleIleIle 446
316 ..GCGCATCATATATGTGAGCGTGCACATACGCGGCTATAAAC 363
|||||:|||||: ||| ||| ||| |||
446 spGlyAspAsnAlaThrIleLys.....AsnThrGlyThrSerAsp 459
364 GTTGAT...TTTGTGCGGAAGAGAAATCCGATCAGACCGTTTTC 410
|||||:|||||: ||| ||| ||| |||
460 IleSerGlyAlaGlySerThrGlyThr..... 468
411 TTACCAATTGTGAAGAAATAATTAATTAAGCTGACATTCACCCCTT 460
::: ||||| |||||: ||| ||| ||| |||
469 .....ValIleAspGlyAsnAsnAlaArgValAsnAsn..... 479
461 ACAAGCGGATTAATCATATGCGCGTTTGCAATAATTGTCACAGATGCA 510
::: ||||| |||||: ||| ||| ||| |||
480 ..AspGlyAsp.....MetThrIleThrAspGly 488
511 GAACCTGCGCAATGACGAGTACATGAGGAGGGAATCTATTCCGATTA 560
488 ..... 488
561 AGAAATATCCGAGCGGTGCGCATGCGCTAGACACCACTATTGGC 610
|||||:|||||: ||| ||| ||| |||
489 .....GlyThrGlyGlyHisIle...T 495
611 GTTATGATGAT.....GACAAACAGCGCATTTATCTACTCTCGGC 651
|||||:|||||: ||| ||| ||| |||
495 hrGlyAspAsnValValIleAspAsnAlaGlySerThrThrValSerGly 511

```



```

3483 GCAGCGCAGCTGATNAGCGGTTATCCCAATAGCGGTTG..... 3522
      ||| :||: |||: |||: |||
1620 .....LeuAspAlaGlyThrThrAsnAsnGluLeuThrSerL 1633
3523 .....AGTGAATTTCCCGCAGCGCTCAGACGCGTTTTC 3555
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1633 euAsnValGlyThrThrAlaGluLeuAsnSerAlaLeuLysGlnValSer 1649
3556 GCCGACAGAGAGATGGACCGCGCTGTTCCGAGACGCC..... 3597
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1650 GtAlaGln.....AlaThrValPheArgGlnAlaArgValLeuSe 1664
3597 ..... 3597
1664 rAsnArgPheThrMetLeuAlaAspAlaAlaProGlnIleLysAspGlyL 1681
3598 .....CGCAACCGCTTTGGACACAGCAGCAGCGGACCAACACACTAC 3642
      |||: |||: |||: |||: |||: |||: |||: |||: |||
1681 euAlaPheAsnValAlaAlaLysGlyAspProArgAlaGluLeuGlyAsn 1697
3643 CGTTGGCAA...GATTCCGGCGCTACCGCCACAAACCGACTCGGC.. 3687
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1698 AspThrGlnTyAspMetLeuAlaLeuArgGlnThrLeuAspLeuThrAl 1714
3688 .....CAATCGGTATGCAGAAAAAACCCTCGGACGCG 3718
1714 aSerLlnAsnLeuThrLeuGluTyGlyIleAlaIArgLeuAspGlyAspG 1731
3719 GGCGCGTGGCATCTCTGTTTCGACAAACGAGCGAAACACTTCGAC 3768
      ||| :||: |||: |||: |||: |||: |||: |||: |||
1731 ly.....SerLyThrAlaGlyAspAsnGlyLeuThr 1741
3769 GACGCGCATCGCACTCGGACGCGTTCCACGCGCGCTTTCGGGCA 3818
      ||| :||: |||: |||: |||: |||: |||: |||: |||
1742 GlyGlyTySerGlnPhePheGlyLeuLysHisSerMetAlaPheAspG1 1758
3819 ATACCGCATCGCG.....AGGTCGACATC...GGCATCA 3850
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1758 u...GlyLeuAlaTrpAsnAsnSerLeuArgTyAspAlaHisAsnLeuA 1774
3851 GCACGCGCGCGGCTTTAGCAGCGCAGTCCTNTCAGACGCGATCGGAGC 3900
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1774 spSerSerArgSerValAlaTyGlyAspValAsnLysIleAlaAspSer 1790
3901 AAAATCGCGCGCGCTGCTGATAC.....GGCATTCAGGCGC 3941
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1791 AspMetArgGlnGlnTyTrLeuGluPheArgSerGluAlaLysThrPh 1807
3942 ATACCGCGCGGCTTTCGCGGATTCGCGATCGAACCGTACATCGGCGAA 3991
      ||| :||: |||: |||: |||: |||: |||: |||: |||
1807 eThrMetMetGly..AspAlaLeuLysValThrProTyArgAlaGlyValL 1823
3992 CGCGCTATTTGCTCCAAAAGCGATTCGCCCTACGAAAACGTAATATC 4041
      ||| :||: |||: |||: |||: |||: |||: |||: |||
1823 ys.....PheArgHisThrMetGluAspGlyTyTrLysGluArgSerAla 1837
4042 GGCACCCCGCGCTGCTGCTCAGCGATACGCGN..... 4074
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1838 GlyAspPheAsnLeuSerMetAsnSerGlyAsnGluThrAlaValAspSe 1854
4075 .....CGGGGCAATTAGCGAGATTTATTCATCAACCGCGCGACACATNT 4120
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
1854 rIleValGlyLeuLysLeuAspTyArgAlaGlyAspLysTrpSerAlaTr 1871
4121 CCATCACN.....CCTATTNTACCGCTGCTCCTATACGAGATGCC 4158
      ||| :||: |||: |||: |||: |||: |||: |||: |||
1871 hrAlaThrLeuGluGlyGlyPro.....AsnLeuSerTy..... 1882
4159 GCTTCGGGCAAGTCCGACACGCGCTCAATACCGCGTATGCGTCAGGA 4208
      ||| :||: |||: |||: |||: |||: |||: |||: |||
1883 ...SerLysSerGlnArgThrAlaSerLeuGlnGlyAlaAlaGlyGlnSe 1898

```

```

4209 TTTCGGC 4215
      :|||: |||
1898 rpheGly 1900

seq_name: SwissProt_40:YFAL_ECOLI

seq_documentation_block:
ID YFAL_ECOLI STANDARD: PRT: 1250 AA-
AC P45508; P45507; P45506; P39441; P76468; P77487;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yfal precursor.
GN YFAL OR B223.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
RA Oshima T., Oyama S., Saito N., Saito Y., Sato Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara C.,
RA Yamagata S., Horinouchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT -K12 genome corresponding to 50.0-66.8 min on the linkage map and
RT analysis of its sequence features."
RL DNA Res. 4:91-113(1997).
RN [3]
RP PRELIMINARY SEQUENCE OF 1-938 FROM N.A.
RX MEDLINE=84272624; PubMed=6087316;
RA Carlson J., Fuchs J.A., Messing J.;
RT "Primary structure of the Escherichia coli ribonucleoside diphosphate
RT reductase operon."
RL Proc. Natl. Acad. Sci. U.S.A. 81:4294-4297(1984).
RN [4]
RP SEQUENCE OF 925-1198 FROM N.A.
RC STRAIN=K12 / EMG2;
RA Estep P., O'Keefe T., Robison K., Church G.M.;
RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE OF 1180-1250 FROM N.A.
RC STRAIN=OV6;
RX MEDLINE=88201664; PubMed=2834621;
RA Hussain K., Elliott E.U., Salmond G.P.C.;
RT "The complete sequence of Escherichia coli also carries a gyrA mutation.
RL Mol. Microbiol. 1:259-273(1987).
RN [6]
RP IDENTIFICATION
RX MEDLINE=96032851; PubMed=7567469;
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RA Danchin A.;
RT "Detection of new genes in a bacterial genome using Markov models for
RT three gene classes."
RL Nucleic Acids Res. 23:3554-3562(1995).
CC -i- SIMILARITY: TO E.COLI TYDEK.
CC -i- CAUTION: REF.3 SEQUENCE DIFFERS EXTENSIVELY FROM THAT SHOWN BY
CC MANY FRAMESHIFTS.

```


418 nllaasplagiylasvalthrvalasnlglyserphealaglyi 435
1721 tcaatagactttaattacagaaataatgcttacacgggttggtt... 1767
435 leileuglialaglylnleuthrillealaglnsnlglyserlyval 451
1768ggcgagaaa.....gatagaccmaaaacga 1793
452 leualaglyalaglnsermetaleuthrglyaspillevalaspas 468
1794 cggcgagctcaactgtttttaccaccccgcgagacacggcaccgc 1843
468 polyalaval...leuserleugllyaspalaalaspheuthral 484
1844 tgccttcggcgagacaat.....ttaaaggcaacatcacggaaca 1887
484 euglnasprproglnserllevalleuasnglylvalleuaspleu 500
1888 AACGGCAAACTGTTTTTCAGCGGACGACGCCGACGCCCTACATCA 1937
501 Serasp.....PheSerTh 505
1938 TTAGAGAACGGGTGCTCAAAAATGGAAGTATC.....CCAC 1975
505 tTprlnserglythrserlyrasnaspilyleuylvalserglyser 522
1976 AAGGAAATCGTGTGGACACGACGTGATCNACCAGCTTTTAACGG 2025
522 erglythrvalileglyserglnasp...valvalaspheualaglyly 537
2026 GAAATTTCCATATTCAGGC..... 2046
538 Aspnasleuhsileglyaspilyllyasprglyvaltyrvalvalva 554
2047GGCAGGCGGTATTTCCGCAATGTGCCAAAGTGG 2083
554 laspalaserspilylvalserleualasnasnserlytleug 571
2084 AA..... 2085
571 lythrthrclnlealaserglythrleumetvalseraspasnerln 587
2086 ...GGGATTTNCATTTAGCAATCAGCCCAAGGATTTTGTGTGCG 2132
588 leuglyasprthrlnstyrasnarg.....glnvalilepethrasply 602
2133 ACCGATCAAC.....CATCAATCTGTACAGTTCGG 2167
602 sglnlnglnservalmetglnlethrseraspvalaspthrlygsera 619
2168 ACTGACNGTCTGACAATTTGTGCAANAANACATTACCGAGATAA 2217
619 spalalaglylnisglyaraspilleglumetargalaspilylval 635
2218 GTGATGTTCTCATTTAGACAGNACNTNAGCGCANTGTNAGCTNNC 2267
636 Alavalaspalaglylvalasprthrlntrpilyalaleumet..... 649
2268 CATATNACGNTNTTNAANCNCNGGCGCTGCNNACTNANAGCATC 2317
650Alasprserglylnh 656
2318 TTAGTGAATGCGATACAGCTTTATAGTCAGCCACAGCGCCCA 2367
656 lsglnaspelyllyserthr.....leuthrlys 665
2368 AAGCGCAACCTTAGCCTGTGGCAATGCCCAAGCAATTTAATCAAG 2417
666 thrgly.....Alaglythrleuglnleuthralaserglyth 678
2418 CACATTAACGGCAACNACATCGNNTCGGCAATGCTTCATTTATCTAA 2467
678 thrlnserlavalargvalgllyllythrleuylsglyaspvala 695
2468 GCAACANCGCCGCAAAAACGGCAGTGTGACGCTTTCGACACGCTAG 2517
695 laasplleuuprotyralasrserleuthrvalgllyaspilylatrn 711
2518GCMAAGTMAAGCATTCGCACTC.....AA 2543
712 Phevalthrlylaspilaspilnserlleasplalileserse 728
2544 CGGCAATGTCCTCCCTAGCCGATTAAGCAGATTCATTTGAAACAGCC 2593
728 rgllythrleasprilleaseraspilythrvalleu.....A 740
2594 GCTTACCGGCAACTGCGCGGACGAGANAGACATTAACCTTAA 2643
740 rgleuthrlyglnaspthrservalaleuasnla...Serleuphe 755
2644 GACAGCAATGACGCTGCGTGACGACGAA.....TTAGG 2681
756 AsnlyaspilyllythrleuvalasnlaThraspilyllythrleuthrly 772
2682 CAATTTAAC.....CTTGACACAG 2701
772 ygluleasnthrasnleugluthraspsrleuthrtyrleuserasn 789
2702 CCACCATTAACATCAATTCGCTATGCGCACGATGCGGCGCA 2751
789 althrvalasnnglyasnleuthrasnthserglyalavalserleugln 805
2752 ACCGCAAGNCTGCAGACAGCGCGCGCGCTTCCGCGCTTCCTATT 2801
806 Asnlylvalalaglyasp..... 811
2802 ATCGGTACACCGCCCACTTCGTAATCCGTTTCAACGCTGACG 2851
812Threuthrv 815
2852 TAAACGCAATTTGAACNGTCAAGAACATTCGCTTATGTGGAATC 2901
815 alasnnglyasprlythrclglyllythrleuuleuaspsrlyleu 831
2902 TTCGCG.....TACGAGCGCAAAATTTAGCTGGCGGAAGTTCCGA 2945
832 Asnlyaspaspservalaseraspilnleuvalmetasnlysnthral 848
2946 AGNACTTACACCTTGGCGGTCAACAT...ACCGC...AAGAACCCG 2989
848 aglyasnthrthrvalavalasnserllethrnglyllyglulpr 865
2990 TAAAGCTCGATCAATTTGACGCTAGTGAAGGAACAAACGCTG 3039
865 hrserthrnglyllylvalvalasphealalaspilaspilprthrlnphe 881
3040 TCGGAAACCTTAATTTACCTTCGCAAAACGAA...CACGTGATGCGG 3086
882 Glnasnasnlaaglnpaserleualaglyserglytyrvalasnmetg 898
3087 CGCGTGGCGTTACCACTCATTCGCAAGACGCGGATTCGCGCTGATA 3136
898 ylatyrasprlythrleuvalgluaspasasnaspilprlyleu... 913
3137 ATCCGTCMAAAGAACAGACTTTCGACAAACTCGGCAAGGAGAAC 3186
913 913
3187 AAAAAACAGCGGAAAAAGACACGCGCAAGCCTTGACGCTGATG 3236
913 913
3237 GCGCGGCGGATGCCCGCAAAAGACAGAAAGCTTCCGCAACGCGCC 3286
914Argserglnlylvalthrproser. 922

```

3287 GCGNCGCAGCGCGGAAAAATGCGCATTTACAGCGGAGAGAGAGAAA 3336
922 .....
3337 AACGGGTGCGAGCGGATAAGACAGCGCNTTGGCGAAACGCGGAGAC 3386
923 .....ProProasProasProth.....
3387 GGAACCGCGCGGNTACACGCGCTTCCCGCGCGCGCGCGCGCGCC 3436
929 .....
3437 GGGATTTGCCGACCGCAGCCGACCGCAACCTCAGCCGCGGAG 3486
930 .....ProasProasProthProasProasProasProthProas 943
3487 CGGACCTATNAGCCGTATGCAATACGGTTGAGTAATTTCCGC 3536
944 ProGluProthProAlaTyrGlnProValLeuAsnAlaLysValGly 960
3537 CACGCTCAACAGCGTTTTCGCGGTACAGAGCAATTCAGCGGTTTG 3586
960 yTyrLeuAsnAsnLeuArgAlaAla.....AsnGlnAlaPheM 973
3587 CCGAAGACCGCGCGCAAC.....GCGTTTGACA 3615
973 etMetGluArgArgAsnHisAlaGlyLysGlyGlnThrLeuAsnLeu 989
3616 AGNCGATCCGCGNACCAACACTACCTCGCAAGATTTCGCGGCTTA 3665
990 ArgValLleGlyLysPThrHisTyrThrAlaLysGlnLeuAla.. 1005
3666 CGGCCCAACAAACCGACCTCGCCAACTCGTATCGAAGAAACCTCGCA 3715
1006 ....GlnHisGluAspThrSerThrValGlnLeuSerGlyAspLeuPheS 1021
3716 GCGGCGGC.....GTCGCGACTCTG..... 3735
1021 etGluArgTyrGlnThrAspGlyGlnTyrMetLeuGlyLysValGly 1037
3736 TTTTGCACACCGCGGCAAAACNCTCGACGAGCGGATTCGCGACAC 3785
1038 TyrSerAspAsnGlnLysPserArgSerAsnMetThrGlyThrArgAl 1054
3786 GGCACGCTTGGCCACGCGCGCTTTTCGGCAATACGCGCATCGCGAGT 3835
1054 aaAspAsnGlnAsnHisGly.....TyrAlaValGly.... 1064
3836 TCGACATCGGCATCAGACGCGCGGCTTTAGACAGCGCGCANCTNTCA 3885
1065 .....LeuThrSerSerThrPheGlnHisGlyAsnGlnLys 1076
3886 GAGCGC.....ATCGGAGCGCA 3902
1077 GlnGlyAlaTyrLeuAspSerThrLeuGlnTyrAlaTyrPheSerAsnAs 1093
3903 AATCCGCGCGCGC.....GTGCTGCATTAC.....GGCATTTC 3934
1093 pValSerGlnGlnLysPglyThrAspHisTyrHisSerSerGlyLleI 1110
3935 AGGCACGATACCGCGCGGTTTC.....GCGGATTTCGCG 3969
1110 laAlaSerLeuGlnLysGlnTyrGlnTyrLeuProGlyArgLysValVal 1126
3970 ATCGAAGCTTACATCGCGCAACGCGCTATTTCGTCGCAAAAGCGGATTA 4019
1127 lIleGluProGlnAlaGlnValLleTyrGlnGlyValGlnGlnAspAspPh 1143
4020 CCGCTACGAAAAAC.....GTCAATTTCG 4042
1143 etThrAlaAlaAsnArgAlaArgValSerGlnSerGlnGlnLysAspLleG 1160
4043 CCAACCCCGGCTTTCGTTCAACGNTACCGNGCGGCGCATTAAGCAGAT 4092

```

```

1160 InThr.....ArgLeuGlyLeuHisSerGlu 1168
4093 TATTCATTCAAACCGCGCAACACATNTCCATCAGCNCCTATTTCAGCT 4142
1169 Trp.....ArgThrAlaValHis.....ValLeuProThrLeuAspLe 1181
4143 GTCC...TATACCGATGCGCGTTCGGCGCAAGTCCGAAACGCGTCATA 4189
1181 uAsnTyrTyrHisAspProHisSerThrGlnLleGlnLysPglySerT 1198
4190 CCGCGNATTTGGCTCAGAGATTTCGCAAAACCGC...AGTGGGAATGG 4236
1198 hr.....IleSerAspAspAlaValLysGlnArgGlyGlnLysVal 1212
4237 GCGCTAAACGCGCGCAATC...AAGGTTTACGCTGTCNTCCAGCGCTGC 4283
1213 GLyValThrGlyAsnLleSerGlnArgValSerLeuArgLysValAl 1229
4284 CGCGCGCAAGAGNCCGCACTGGAAGCGCAACAGCGCGGCGCATCAAT 4333
1229 aTrrGlnLysGlySerAspAspPheAlaGlnThrAlaGlyPheLeuSerM 1246
4334 TAGGCTACCGCTGG 4347
1246 etThrValLysTrp 1250

seq_name: SwissProt_40:YEJO_ECOLI
seq_documentation_block:
ID YEJO_ECOLI STANDARD; PRT; 863 AA.
AC P33924; P76450; Q47291;
DT 01-FEB-1994 (rel. 28, Created)
DT 01-FEB-1994 (rel. 28, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Hypothetical outer membrane protein yeJo.
GN YEJO OR B2190.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / BHB2600;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474(1997).
RN [3]
RP PRESENCE OF AN INSERTION SEQUENCE.
RA Rudd K.B.;
RL Unpublished observations (JAN-1994).
CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -1- SIMILARITY: SPRONG, TO BORETELLA PERTACTIN.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE READING FRAME IS
CC INTERRUPTED BETWEEN CODONS 21 AND 22 BY A 15K INSERTION ELEMENT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
```



```
2687 TAAACCTTGACAACGCCACCATTTACATCAATTCGCCCTATTCGCACGAT 2736
      |||  ::  ::  ::|||::|||::|||::|||
      1ethrThrThrAsnGlnAsnThrLeuLeuAsnLeuAla.....AsnLeu 433
2737 GGTGACAGCGCGCAACCGGCAGNGTGTGACAGACCGCGCGCGCGCTTC 2786
      |||  ::  ::  ::|||::|||::|||::|||
      434 AlameSerAspAlaAsnValIleMetMetAspGluProValThrThrGse 450
2787 GCGCGCGTTCCTATTCCTGCTTACACCGCCAACTTCGTAGAAATCCCTT 2836
      |  |||||::|  ::|||::|||::|||::|||
      450 r.....SerValThr.....AlaSerAlaGluAsn...P 459
2837 TCACACGCTGACGAGTAAACGGCAATTTGACNGTCAAGAACATTCGCG 2886
      ||  |||||::|||::|||::|||::|||
      459 heileThrLeuThrThrAsn...ThrLeuSerGlyAsnGlyAsnProLeu 474
2887 TTTATGTCGGAACCTTTCGGCTACGAGACGCAAAATTTGAAGTGGCGGA 2936
      ::  ::|||::|||::|||::|||::|||
      475 MetArgThrAspMetAlaAsnHisGlnSerAspGlnLeuAsnValThrG1 491
2937 AAGTTCGGAAGNACTTACACCTTGGCGGTCAACAATACGGCAACGAGAC 2986
      ::|||::|||::|||::|||::|||
      491 yGlnAlaThrGlyAspPheGlyIlePheValThrAspThrGlyAlaSerP 508
2987 CCTTAAGCCTCGATCAATTGACGCGTAGTGGA.....GGGAAGAACAAC 3030
      ||::|||::|||::|||::|||
      508 rolaIaIaGlyAspSerLeuThrLeuValThrThrGlyGlyAsp... 523
3031 AAACCGCTGTCCGAACCTTAATTTCACCTTCGCAAAAGCAACAC..... 3075
      |||||::|||
      524 .....AlaAlaPheThrLeuGlyAsnAlaGlyGlyA 534
3076 .GTCAATGCCGCGCGCTGGCTTACCACTCATCGCAAAAGACGCGAGT 3124
      |||||::|||::|||::|||::|||
      534 lValaSpIleGlyThrGlyGlyThrLeuLeu..... 545
3125 TCCGCTGCATATCCGTCAAAGAACAGACGCTTCCGACAACTCGGC 3174
      |||||::|||
      545 ..... 545
3175 AAGGCAGAGCCAAAMAAACAGCGGAAAGAACAGCGCAAAAGCCTTGA 3224
      |||||::|||
      545 ..... 545
3225 CCGCGCTGATTCGCGCGCGCGGATTCGCGCGCAAAAGACGAGACGTTG 3274
      |||  ::|||::|||::|||
      546 .....AspAsnGlyAsnHisSerThrSerLeuA 555
3275 CCGAACCGCGCGCGCGGAGCGGGAATGTGCGCATTTATGACAGCG 3324
      |||||
      555 lAgIu..... 556
3325 GAGGAAGAGAAAAACGGGTGACGCGGATTAAGACAGCGCNTTGGCGAA 3374
      |||||::|||
      556 ..... 556
3375 ACAGCGCGAAGGGAACCGCGCGGATTCACACGCGCTTCCCGCGCGCC 3424
      ::|||::|||::|||::|||
      557 ...AsnAlaGlyAlaGlnIleThrProSerThrThr..... 566
3425 GCNGCGCGCGCGGATTTGCCGCAACCGCACGCCCAACGCAACCTCAA 3474
      |||||
      566 ..... 566
3475 CCCCACACGCGACCGGACCTGTATNAGCGGTATGCAATAGCGGTTGAG 3524
      |||||::|||
      567 .....AspValLeuAsnMetAlaAlaAlaGlnProLeu... 577
3525 TGAATTTCCGCGACGCTCAACAGCGTTTTCGCGTACAGAGCAAGTGG 3574
      |||||::|||
      578 .ValPheAspAlaGlnLeuAspThr.....ValArgGluArgLeuG 591
3575 ACCGCGTGTTCGCGAAGACCGCGCAACGCGNGTTTGGCAAGGCGCATC 3624
      |||||::|||
      591 lSerValLeuSerGlyValSerThrAspThrAlaMetThrSerSerAlaIle 607
3625 CGGNAACACCAACACTACCGCTTCGCAAGATTTCCGCGCTTACCGCAACA 3674
      ::|||::|||::|||::|||
      608 AsnThrArgAsnAsnValThrThrAspAlaGlyAlaGlyPheGluGlnTh 624
3675 AACCGACTGGCGCAATCGGTATGCGAAGAAACCTC.....G 3712
      |||||::|||
      624 rLeuThrGlyLeuThrLeuGlyIleAspSerArgPheSerArgGluGln 641
3713 GCAGCGGCGCGCGCATCCCTGTTTTCGCAACCGGACGCAAGAACANC 3762
      ::|||::|||::|||::|||
      641 erSerThrIleArgGlyLeuLeuPheGlyIleSerHisSerAspIleGly 657
3763 TTCGACGACGCGCATTCGGCACTCGGACGCGCTTCCACGCGCGCTTT 3812
      |||||
      658 PheAspArgGly..... 661
3813 CCGGCAATACGCAATCGGACGCTTCGACATCGCATCAGACGCGCGCGG 3862
      |||||::|||
      662 .GlyIleGlyAsnIleAspSerThrThrLeuGlyAlaThrAlaGlyTrpG 678
3863 GTTTTACGCGCGCATCTTTCAGACGCGC..... 3891
      ::|||::|||
      678 luhIsgIuhAsnGlyAlaThrValAspGlyValAlaValysValAspArgPhe 694
3892 .....ATCGGACGCAAAATCCGCGCGCGCTGCTGATTCAGGC.. 3930
      |||||::|||
      695 AlaAsnThrIleHisGlyIleGlyMetSerAsnGlyAlaThrAlaPheGlyAs 711
3931 .....ATTCAGGACGACGATTCAGCGCGCG 3952
      ::|||::|||
      711 pTyraSerAsnGlyAlaGlyAlaHisValGlnSerGlyPheArgTrpY 728
3953 GTTTCGCGGATTCGCGCATCGAACCGTACATCGCGGACGCGCCTATTTG 4002
      |||  ::|||::|||::|||
      728 alaSpGlyLeuThrPserValArgProThrLeuAlaPheThrGlyPheThr 744
4003 GTCCAAAAGCGGATTCACCGCTACGGAACGTCATTCGCCGCCCGCG 4052
      ::|||::|||
      745 ThrAspGlyGlnAspGlyThrLeuSerAsnGlyMetArgAlaAspValG1 761
4053 T...CTTGCCTCAACCGNTACCGGCGGCGCATTAAGCGATTTATTCAT 4099
      |  ::|||::|||::|||
      761 yAsnThrArgIleLeuArgAlaGlnAlaGlyThrAlaValSerThrHisM 778
4100 TCNAACCGCGCGCAACATTCATTCACNCCTTATTTAGCCTGCTCTAT 4149
      ::|||::|||
      778 etAspLeuGlnAsnGlyThrThrLeuGlnProThrProLeuLysAlaAlaVal 794
4150 ACCGATGCGCGCTTCGCGCAAGTCCGCAACGCGGATTCATACCGCNCNTAT 4199
      ::|||::|||
      795 ArgGlnIleuThrAlaAspSerAsnGlnValLysValAsn..... 807
4200 GCGTCAGGATTCGCGCAAAACCGCACTGCGGATGCGCGGCTTAAC.... 4245
      |||  ::|||::|||::|||
      808 .....AspAspGlyLysPheAsnAsnAspValAlaGlyThrSerGlyY 822
4246 .....GCCGAATCAAAAGTTTC.....ACGCTGTCNCNTCAC 4278
      |||  ::|||::|||::|||
      822 a1TyrglnAlaGlyIleArgSerPheThrProThrLeuSerGlyHis 838
4279 GGTGCGCGCGCGCAAGNCGCAACTGGAAGCGCAACGCGCGCGCGCAT 4328
      ::|||::|||
      839 LeuSerValSerThrGly.....AsnGlyAlaGlyA 849
4329 CAAA 4332
      ::|||
      849 lGlu 850
seq_name: SwissProt_40:YDUA_ECOLI
```

```

seq.documentation_block:
ID YDBA_ECOLI STANDARD; PRT; 2003 AA.
AC P33666; P76087; P76088; P76856; P76857; P76859;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ydbA.
GN YDBA OR B1401/B1405.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alpa H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasundaram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 464-2003 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92190338; PubMed=1665988;
RA Moszer I., Glaser P., Danchin A.;
RT "Multiple insertion sequences near the replication terminus in
Escherichia coli K-12.";
RL Biochimie 73:1361-1374(1991).
CC -1 SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYS6 (AC P25928).
CC -1 CAUTION: THIS IS A CONCEPTUAL TRANSLATION: THE GENE CODING FOR
THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT
BETWEEN AMINO ACIDS 839 AND 840.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000237; AAC74483.1; ALT_SEQ.
DR EMBL: AE000237; AAC74487.1; ALT_SEQ.
DR EMBL: D90778; BAI15009.1; ALT_SEQ.
DR EMBL: D90778; BAI18880.1; ALT_SEQ.
DR EMBL: D90779; BAI18881.1; ALT_SEQ.
DR EMBL: X62680; -; NOT_ANNOTATED_CDS.
DR Ecogene: EG1307; ydbA.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 489 489 I -> V (IN REF. 2).
FT CONFLICT 495 495 I -> V (IN REF. 2).
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;

```

```

alignment_scores:
Quality: 242.00 Length: 1528
Ratio: 0.358 Gaps: 77
Percent Similarity: 44.241 Percent Identity: 19.241

```

```

alignment_block:
US-09-303-518d-651 x YDBA_ECOLI ..
Align seg 1/1 to: YDBA_ECOLI from: 1 to: 2003
82 TTAGCCATATGCGTGTGGGCAATTCCTCCCAAGCTTGGCGGAC 131
7 LeuSerAlaCysIleAlaLeuAlaLeuSerGlyGlnGlyTrpAlaAla... 22
132 CACTTATTTCGGCATCACTACCAATCTATCCGACATTGCCGAATA 181
23 .....AspIleThrGlu.... 26
182 AAGCAGATTTCAGTCGGCGCAA.....GATATTGAGCTTAC 222
27 ..ValGluThrThrGlyGlyLysLysAsnThrAsnValThrCysPro 42
223 AACAAAAAGGAGGAGTTGGTGGCAATCAATGACAAAGCCCGATAT 272
43 AlaSprProGlyLysLeuSerProGluGluLeuLysArgLeuProSerGI 59
273 TGATTTTCTGTGCTGTCGCGTAAC.....GGCG 301
59 uCysSerProLeuValGluGlnAsnLeuMetProTrpLeuSerThrGlyA 76
302 TGGCGGATGTTGGGCGGATCAATATATGTGTGAGCGTGGACATAACGGC 351
76 IAlaIAlaLeuIleThrAlaLeuAlaValAlaGluLeuAsnAspAsp 92
352 GCGTAT.....AACACGTTGATT..... 372
93 AspHisHisIAsArgAsnAsnSerProLeuProThrProProAspAs 109
373 .....GTCGCGAAGAA 388
109 pGluSerAspAspThrProValProProThrProGlyAspGluIleI 126
389 ATCCCGATTCACACCGTTTCTTACCAATGTGAAAGAAATATAT 438
126 IeProAspAsp..... 129
439 AAGCTGACATTCACAC.....CCTTACACGGCA 470
130 ...ProAspAspThrProThrProProLysProValSerPheAsnAs 145
471 TTACCATATGCGCGTTGCATAATTGTGCACGATGACGACCTTCG 520
145 pValIleLeuAspLysThrGluLysThrLeuThrIleArgAspSerValp 162
521 AAATGACGAGTACATGAGGGGAATACCTATCCGATAAAGAAATAT 570
162 herThrTyThrGluAsnAlaAspGlyThrIleSerLeuGlnAspSerAsn 178
571 CCCGAGGTGCCCATGCGCTCAGACACCATTCATTCATGATGAT 620
179 GlyArgLysAlaThrIle.....AsnLeuTrpGlnIle...As 190
621 TGACAAACACGGCGATTTATCTTACTCCGGCGCATGTATATGCGGCA 670
190 pGluIleAsnAsnThrValAlaLeuGlnGlyVal...SerAlaAspLys 206
671 ATACACATATGCGGTTGGGGAATATAGCCGTANTTATGAGCGGC 720
206 IatThrLysTrpGln...TyrAsnHisAsnGlyGluLeuValIleThrGly 221
721 GATGTGGCGCATGCAACGACATATGCGCTATGCGGATGCGAGTGGCGC 770
222 AspAsnIatThrValAsnAsnGlyLysThrThrValAspGlyLys.. 237
771 AGCGACACGCGTTCGCATGTTATTATGACAAACAAACATTAAT 820
238 .....AspSerThrGlyThrGlu.. 243

```

```
821 GGCTGCTCAACGA.....GTTTACAAACCGGC...TACCT 855
      ::::|
      ||::|
244 .....IleasnGlyAsnAsnGlyLysValIleGlnAspGlyAspLeuAsp 258
      ::::|
      ||::|
856 TATTCGGCGGAGGAAACGGTTTCCAGCTGATACGCAAAAGATTGGTTCA 905
      ||::|
      ::::|
259 ValSerGlyGlyLysIleAspIle..... 268
      ::::|
      ||::|
906 CGATGACATTACAGAGCGATACACATCCGCTCTTTTGAACCGCGCA 955
      ||::|
      ::::|
269 .....ThrGlyAspSerAlaThrVal..... 275
      ::::|
      ||::|
956 GTAACGACATTTTTCCTTACATCCAAACACACGCTACGGGTACGTA 1005
      ::::|
      ||::|
276 .....AspAsnGlySerThrMetThrVal 283
      ::::|
      ||::|
1006 ACAGAAACACAGAAAGGNTCCATCCAAAGCTTAAAGTACACAGCT 1055
      ||::|
      ::::|
284 ThrAspProGlu.....SerMetGlyIleGlnIleAs 294
      ::::|
      ||::|
1056 CCGACTGTTGACGATCTTGAATGAATGATAAGTAAACACAGTTAGC 1105
      ||::|
      ::::|
294 PGL.....AspLysAlaIleValAsnAsnGlyGlyLysSerThrIleT 309
      ::::|
      ||::|
1106 CGCGAGGGGGGTAAATCACTAC.....CGTCCAAAGTTA 1140
      ::::|
      ||::|
309 hrAsnGlyGlyThrGlyThrGlnIleAsnGlyAspAspAlaThrAlaAsn 325
      ::::|
      ||::|
1141 AACACGGTGAACCT.....TCCTTTAT 1166
      ||::|
      ::::|
326 AsnAsnGlyLysThrThrValAspGlyLysAspSerThrGlyThrVal 342
      ::::|
      ||::|
1167 CGATTACGGCAACGCAATCATCTTATCAACAACATCAACCAAGCGC 1216
      ::::|
      ||::|
342 eaSnGlyAsnAsnGlyLysValIleGlnAspGlyAspLeuAspValSerG 359
      ::::|
      ||::|
1217 CGGGC.....GGTTGATTTTGAAGTATTTACGGTCTCGCTGAA 1260
      ::::|
      ||::|
359 LysGlyLysIleAsnGlyIleAspIleThrGlyAsp...SerAlaThrValAsp 374
      ::::|
      ||::|
1261 AACACGAACAGCTG.....CAAGGGCGGGGGCTCATAT 1295
      ||::|
      ::::|
375 AsnLysGlyThrMetThrValThrAspProGluSerIleGlyIleGlnVa 391
      ::::|
      ||::|
1296 CAGTGAAGACAGTACCGTACTTGAAGTAAACGCGCTGCGCAACGACC 1345
      ::::|
      ||::|
391 LAspGlyAspGlnAlaVal.....ValAsnAsnGlyGlyLysSerA 405
      ::::|
      ||::|
1346 GCCTGTCCAAATCGCAAAAGCAGCCTGACGTTCAAGCCAAAGGGGAA 1395
      ::::|
      ||::|
405 lAlaThrAsnGlyGlyThrGlyThr.....GlnIleAsnGlyAsp 418
      ::::|
      ||::|
1396 AACCAAGCGCTCAGTCAGCGTGGCGCAGCGTACAGTCATTTTGGATCAGA 1445
      ::::|
      ||::|
419 AspAlaThrAlaAsnAsnAsnGlyLysThrThrVal..... 430
      ::::|
      ||::|
1446 GGCAGACGATAAAGCAAAACAGCTTGTAGTAATCGGCTGNTCA 1495
      ::::|
      ||::|
431 .....AspGlyLysAspSerThrGlyThrGlnIleAlaGlyAsnA 444
      ::::|
      ||::|
1496 GCGGCGAGGGGTACGTGCAACTGAATCCGATTAATCAGTTCAACCCCGAC 1545
      ::::|
      ||::|
444 snGlyLys..... 446
      ::::|
      ||::|
1546 AAACCTATTTCGGGCTTCGGCGGAGAGCTTGGATTAAAGCGCATTC 1595
      ::::|
      ||::|
447 .....ValIleGlnAspGlyAspLeuAspValSerGlyGly 459
      ::::|
      ||::|
1596 GCTTTCGTTCCACCGTATTCAAATACGATGAGGGGAGATGATGNCN 1645
      ::::|
      ||::|
459 Y.....HisGlyIleAspIleThrGlyAspSerAlaThrVal.... 471
      ::::|
      ||::|
1646 ATCATTAATGCCACAACAACATCCACCGCTTACCATTAACAGGAATGAAGT 1695
      ::::|
      ||::|
472 .....AspAsnLysGlyThrMetThrValThrAspProGluSer 484
      ::::|
      ||::|
1696 ATT...ACACAACCGAGTGTAAAGAAATTCATTAAGTAAATTAACCAA 1742
      ::::|
      ||::|
485 IleGlyIleGlnIleAspGlyAspGlnAlaIleValAsnAsnGlyGly 501
      ::::|
      ||::|
1743 AGAATTCCTTACCAACGGTGTGGTGGGAGAA.....GATA 1780
      ::::|
      ||::|
501 userThrIleThrAsnGlyGlyThrGlyThrGlnIleAsnGlyAsnAspA 518
      ::::|
      ||::|
1781 CGACCAAAACGACAGCGCGCTCAACCTGTTTACACGCGCCGCGAGAA 1830
      ::::|
      ||::|
518 lAlaThrAlaAsnAsnSerGlyLysThrThrVal.....AspGlyLys 531
      ::::|
      ||::|
1831 GACCGCACCCCGCTGCTCCGCGGA.....ACAATTTAAA 1868
      ::::|
      ||::|
532 AspSerThrGlyThrLysIleAlaGlyAsnIleGlyIleValAlaAsnLeuAs 548
      ::::|
      ||::|
1869 CGGCACATCAACGCAACAAACGCAAACTGTTTTCAGGCGGACAGCA 1918
      ::::|
      ||::|
548 PGLSerLeuThrValThrGlyGly..... 556
      ::::|
      ||::|
1919 CACCGCACCGCTACATCATTTAGAAACGCGGTGTCAAAATGGAAGT 1968
      ::::|
      ||::|
557 ..AlaHisGlyValGlnAsnIleGlyAsp..... 565
      ::::|
      ||::|
1969 ATCCACAGAAGAGAAATTCGTGGGACACACACTGCATCAACCCACGTT 2018
      ::::|
      ||::|
566 .....AsnGlyThrValAsnAsnLysGlyAspIleValAlaSerAspTh 580
      ::::|
      ||::|
2019 TAAAGCGGAATTTCCATATATTCAGGGCGGCGGCGGTGATTTCCCGCA 2068
      ::::|
      ||::|
580 rGlySerIleGlyValLeuIleAsnGlyGlyGlnAlaThrValSerSnt 597
      ::::|
      ||::|
2069 ATGTTGCCAAAGTGAAGCGGATTCATTTGACCAATCACGCCCAAGCA 2118
      ::::|
      ||::|
597 hr.....GlyAspValAsnValSerAsnGlnAlaThrGly 608
      ::::|
      ||::|
2119 GTTTTGGTGTGCGACCGCAT..... 2139
      ::::|
      ||::|
609 ...PheSerIleThrThrAsnSerGlyLysValSerLeuAlaGlySerMe 624
      ::::|
      ||::|
2140 ..CAAGCCATCAATCTGTACACGCTGACGTGACGCGTCTGCAAAAT 2188
      ::::|
      ||::|
624 tGlnValGlyAspPheSerThrGlyValAspLeuAsnGlyAsnAsnSns 641
      ::::|
      ||::|
2189 GTGTGGAANAANCATTACGACGATTAAGTATGCT...TCATTGACT 2235
      ::::|
      ||::|
641 eValThrLeuAlaAlaLysAspLeuLysValAlaGlyGlnLysAlaThr 657
      ::::|
      ||::|
2236 AAGACNCACTNNAGCGCANTGTAAGCTNNCCATTAACGTTNNTTAAA 2285
      ::::|
      ||::|
658 GlyIleAsnValSerGlyAsp..... 664
      ::::|
      ||::|
2286 NCTCNCNGGCGTNCNCACTNNAAGCAATCTTGTGCAATGCGCAT 2335
      ::::|
      ||::|
665 ...AlaAsnThrValAsnIleThrGlyAsnValLeuValAspLysSPL 680
      ::::|
      ||::|
2336 CACGT..... 2340
      ::::|
      ||::|
680 yThrAlaAspAsnAlaAlaGlnIleThrPheAspProSerValGlyIle 696
      ::::|
      ||::|
2341 .....TATACAGTACCCACACACGCCCAAAACGCAACTTACCT 2384
      ::::|
      ||::|
697 AsnValThrGlySerAspAsnAsnValThrLeuAspGlyLysLeuThrVa 713
      ::::|
      ||::|
2385 CGTGCGCAATGCCCAACCAATTAATCAAGCCACTTA...AACGCA 2431
      ::::|
      ||::|
713 lValSerAspSerGlnValThrSerArgGlnSerAsnLeuPheAspLys 730
      ::::|
      ||::|
2432 ACNCA...TCGNTTCGGCAATGCTTCATTTAATCAACCAACAGCC 2478
      ::::|
      ||::|
```



```

seq_name: SwissProt_40:OMPA_RICRI
seq_documentation_block:
ID OMPA_RICRI STANDARD; PRT; 2249 AA.
AC P15921;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface
DE antigen) (OMP A) (omp A).
GN OMPA.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiinae; Rickettsia.
OX NCBI_TaxID=783;
(1)
RP SEQUENCE FROM N.A.
RC STRAIN=R;
RX MEDLINE=90354033; PubMed=2117568;
RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;
RT "A protective protein antigen of Rickettsia rickettsii has tandemly
RT repeated, near-identical sequences."
RL Infect. Immun. 58:2760-2769(1990).
CC -I- FUNCTION: ELICITS PROTECTIVE IMMUNITY.
CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -I- PTM: GLYCOSYLATED (PROBABLE).
CC -I- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31227; AAA26380.1; -.
DR PIR; A41477; A41477.
DR InterPro; IPR003858; OMPA_rOMPb.
DR Pfam; PF02708; ompa_rompb; 1.
KW Antigen; Repeat; Signal; Cell wall; S-layer; glycoprotein.
FT SIGNAL 1 28
FT CHAIN 29 2249 OUTER MEMBRANE PROTEIN A.
FT DOMAIN 212 1180 13 x APPROXIMATE TANDEM REPEATS.
FT REPEAT 212 286
FT REPEAT 287 358
FT REPEAT 359 430
FT REPEAT 431 505
FT REPEAT 506 577
FT REPEAT 578 652
FT REPEAT 653 724
FT REPEAT 725 799
FT REPEAT 800 874
FT REPEAT 875 949
FT REPEAT 950 1021
FT REPEAT 1022 1093
FT REPEAT 1094 1165
FT REPEAT 1166 1180
FT REPEAT 1166 1180 TYPE I (INCOMPLETE).
SQ SEQUENCE 2249 AA; 224333 MW; A9D6646C089DF087 CRC64;

```

Ratio: 0.356 Gaps: 72
Percent Similarity: 44.256 Percent Identity: 19.125

Alignment_block:
US-09-303-518D-651 x OMPA_RICRI ..

Align seg 1/1 to: OMPA_RICRI from: 1 to: 2249

```

566 AATATCCCAAGCGTGTCCGATGGCGCTACGACACCACTATGGC..... 610
      :::::  :::::|||||:::|||||  :::::|||||
880 SerLeuAlaThrIleSerValGlyAlaGlyThrValaThrLeuGlyGlyAl 896
611 .....GTTATGATGATGACAAACAGCGCAT..... 636
      ::  ::  :::::|||||:::
896 aValIleLysAlaThrThrThrLysLeuThrAsnAlaIleSerValLeuT 913
637 ..TTATCTACTCTCCGCGCATGGTAAATGCGCGCAATACACATATGCAG 684
      ||:::  :::::|||||  ||  ||:::  ::
913 hLeuThrAsnAlaAsnAlaValLeuThrGlyAlaIleAspAsnThrThr 929
685 GGTGGGGAAATATATGCGCTAANTACTTGC..... 714
      |||  |||  |||||  ::::|
930 GlyGlyAspAsnValGlyValLeuAsnLeuAsnGlyAlaLeuSerGlnVa 946
715 .AGCGCGATGTGTCCGCATCCACACGATATGGCGCTATGCCGATATGCGAC 763
      ::|||:::  :::::|||||:::  ::::  :::::
946 lThrGlyAspIleGlyAsnThrAsnSerLeuAlaThrIleSerValGlyA 963
764 GTGCGCGACGCGACAGCGTTGCCCAATGTT..... 795
      :::::  :::::  ||:::  :::::
963 lAglyThrAlaThrLeuLeuGlyAlaValIleLysAlaThrThrThrLys 979
796 ATTATGACAAAAACAACATATATGGCTCTCTACAGGAGTTTACAAC 845
      ::::  |||  :::::  :::::  |||  ||:::  ||
980 LeuThrAspAlaIleSerAlaValLysPheThrAsnProValValValTh 996
846 CGCG...TACCCATTATCCGCGAGGAAAAAGTTCCAGCTGATACGCA 892
      ||||  :::::  :::::  :::::  |||
996 rGlyAlaIleAspAsnThrGlyAsnAlaAsn..... 1006
893 AAGATTGTTCTACGATGACATTACAGAGCGCATACATACCGCTGTT 942
1006 ..... 1006
943 TTGGAACCGCGCATACGACATTTTCTTATCATCAACAACAACG 992
      |||||  :::::|||||:::|||||:::
1007 .....AsnGlyIleValThrPheThrGlyAsnSerThrVa 1018
993 TACGGGTACGGTACAGAAACAACGAA..... 1020
      |||||  :::::|||||
1018 lThrGlyAsnValGlyAsnThrAsnAlaLeuAlaThrValAsnValGlyA 1035
1021 .....AAGTNTCCAAATCCAAAGCTTAAAGTACAGACGTCGCA 1059
      ::|||:::  ::|||:::  ::|||:::  ::|||:::
1035 lAglyLeuLeuGlnValGlnGlyGlyValValValAsnThrIleAsn 1051
1060 GTGTTGACGAATCTTTGAATGAAGTATGAACACCACTTTACGCGGC 1109
      |||  |||  |||  ::|||:::  ::|||:::
1052 LeuThrAspAsnAlaSerAlaValThrPheThrAsnProValValValTh 1068
1110 AGCGGCTGTTATCATACCTACCGTCCACAGGTTAAACAACGGGGA... AAC 1156
      ::|||:::  ::|||:::  ::|||:::  ::|||:::
1068 rGlyAlaIleAsp.....AsnThrGlyAsnAlaAsnA 1079
1157 TTTCTTTATTCGATTAACGCGACGCAACATCTTATATCAAC..... 1200
      :::::  ::  ::  :::::  ::  ::  ::|||
1079 snGlyIleValThrPheThrGlyAsnSerThrValThrGlyAsnValGly 1095
1201 .....AACATCAACCAAGCGCGGCGGCTTTGTATT 1222
      :::::||||  |||||  |||  |||
1096 AsnThrAsnAlaLeuAlaThrValAsnValGlyAlaGlyLeuLeuGlnVa 1112
1233 TGAAGTGATTTTACGCTCTCGCTGAAAAACAACGAACGTAACGCAAGCG 1242

```

```
1112 1GInGlyValValLysAlaAsnThrLeuThrAspSnaLas 1129
1283 CGGGCGTTCAATATCAGTGAACAGTACCGCTTACTTGGAAAGTAAAGCC 1332
1129 erAlaValThrPheThrAsnProValValThrGlyAlaIleAspAsn 1145
1333 GTGGCAACGACCGCTGTCCAAATCGGCAAGACGCGTCACGTTCA 1382
1146 ThrGlyAsn..... 1148
1383 AGCCAAAGGGAAACCAAGCGTCGATCAGC...GTGGCGACGCTACAG 1429
1149 .....AlaAsnAsnGlyIleValThrPheThrGlyAsnSerThrV 1162
1430 TCATTTTGATCAGCAGCAGCATTAAGCAAAAAACAAGCCTTAGT 1479
1162 a.....ThrGlyAspIleGlyAsnThrAsnAlaLeuValA 1173
1480 GAAATCGGCTGNTCAGCGCGAGGGTACGGTGCACGTGAATGGCGATA 1529
1174 ThrValAsnValGlyAlaGly.....IleThrLeuGlnAlaGlyG 1187
1530 TCAGTTCAACCCGCAAACTGTATTTCGGCTTCGC..... 1566
1187 YSerLeuAlaAlaAsnAsnIleAspPheGlyAlaArgSerThrLeuGlnP 1204
1567 ..GGCGACGCTTTGGATTTAACGGCATTCGCTTGTCCACCGATT 1614
1204 heAsnGlyProLeuAspGlyGlyIleGlyAlaIleProGlyTyr..... 1218
1615 CAAATACCGATGAGGGCGATGATTCNCATCATATATGCC..... 1656
1219 .....PheGlyGlyAlaIleAlaAsnGlyAsnAsnAlaIleLeuAs 1232
1657 .....ACACACATCCACCGTTACATT..... 1680
1232 nValAsnThrLysLeuLeuThrAlaSerHisLeuThrIleGlyThrValA 1249
1681 .....ACAGGAATGAAGTATTACACACGAGGAGT 1713
1249 IaGluIleAsnIleGlyAlaGlyAsnLeuPheThrIleAspAlaSerVal 1265
1714 AAGATATCAATAGACTTAATTACAGCAAAAGATTGCTTACACGGTTG 1763
1266 GlyAspValThrIleLeuAsn.....AlaGlnAsnIleAs 1277
1764 GTTGGCGAGAAAGATACGACCAAAAGACGCGCGCTCAACCTTGT 1812
1277 nPheArgAlaArgAspSerValLeuValLeuSerAsnLeuThrGlyValG 1294
1813 .....TACCAACCCGCGCGCAAA 1830
1294 lValAlaAsnIleLeuLeuAlaAlaAspLeuValAlaProGlyAlaAsp 1310
1831 GACCGCACCCGCTGCTTTCGCGGGAACA.....AATTAAACGG 1871
1311 GluGlyThrValValPheAsnGlyGlyValAsnGlyLeuAsnValGlySe 1327
1872 CAACATCCGCAACA..... 1887
1327 rAsnValAlaGlyThrAlaArgAsnIleGlyAspGlyGlyAsnLysP 1344
1888 ..AACGGCAACTGTTTTCAGCGGAGACGACGACGCGACGCTTACAT 1935
1344 heAsnThrLeuLeuIleTyrAsnAlaValThrIleThrAspAspValAsn 1360
1936 CATTAGGAACGGGTGCTCAAAATGAGAGTATCCACAGAGAGAAAT 1985
1361 .....LeuGlnGlyIleGlnAsnValIleuII 1369
1986 CGTGGGACACGACTGATCMAACGACG...TTTAAACGGCAAAAT 2032
I.....T
```

```
1369 eAsnLysAsnAlaAspPheThrSerSerThrAlaPheAsnAlaGlyAlaI 1386
2033 TCCATATTCAGGGCGGCGACGCGTATTTCCCGCAATGTCCCAAGG 2082
1386 IeGlnIleAsnAspAlaThrTyrThrIleAspAlaAsnAsnGlyAsnLeu 1402
2083 GAA.....GGCATGNCATTGAGCAATCAACGCCACAGA..... 2118
1403 AsnIleProAlaGlyAsnIleGlnPheAla...HisAlaAspAlaGlnLe 1418
2118 ..... 2118
1418 uValLeuGlnAsnSerSerGlyAsnAspArgThrIleThrLeuGlyAlaA 1435
2118 ..... 2118
1435 snIleAspProAspAsnAspAspGlyIleValIleLeuAsnSerVal 1451
2119 .....GTTTGGTGTGCG 2132
1452 ThrAlaGlyLysLysLeuThrIleAlaGlyGlyLysThrPheGlyGlyAl 1468
2133 ACCGATCAAAAGCCATACATC.....GTACAC 2161
1468 a...HisLysLeuGlnThrIleLeuPheLysGlyAlaGlyAspCysSer 1484
2162 GTTCGGACTGGACN.....GCTCGACAAAT 2187
1484 hrAlaGlyThrThrPheAsnThrThrAsnIleValIleuAspIleThrGly 1500
2188 TGTGTGGAANAANCAATTACCGACGATAAAGTGAT..... 2223
1501 GlnLeuGlnLeuGlyAlaThrThrAlaAsnValValLeuPheAsnAspAl 1517
2224 ..GCTTATTGACTAGACGACGACNTNAGCGC..... 2253
1517 aValGlnLeuThrGlnThrGlyAsnIleGlyGlyPheLeuAspPheAsnA 1534
2253 ..... 2253
1534 lAlAsnGlyMetValThrLeuAsnAsnAsnValAlaGlyAla 1550
2254 .....ANTGNAGCTNCCNA 2270
1551 ValGlnAsnThrGlyGlyThrAsnAsnGlyThrLeuIleValLeuGlyAl 1567
2271 TNACGNTNNTTNAANCTCNCNGCGCNTGONNCACIN.....AANG 2311
1567 aSerAsnLeuAsnArgValAsnGlyIleAlaMetLeuLysValGlyAlaG 1584
2312 GCAATCTTAGTGCAAT.....GGCGATACAGT 2340
1584 lAsnValThrIleAlaLysGlyGlyLysValLysIleGlyGlnIleGln 1600
2341 TATACAGTCAAGCCCAACGCCCAAAAGCAACCTTAGCTGCTGGG 2390
1601 GlyThrGlyThrAsnThrLeuThrLeuProAlaHisPheAsnLeuThrG 1617
2391 CAATGCCCAAGCAACATTTATCAAGCC...ACATTAAAC..... 2427
1617 YSerIleAsnLysThrGlyGlyGlnAlaLeuLysLeuAsnPheMetAsnG 1634
2428 .....GGCAC 2433
1634 lGlySerValSerGlyValValGlyThrAlaAlaAsnSerValGlyAsp 1650
2434 NCATCGNTTCGGCAATGCTTCATTTAATCTAAGCAACAGCC..... 2478
1651 IleThrThrAlaGlyAlaThrSerPheAlaSerSerValAlaAlaLysG 1667
2479 .....GCACAA 2485
1667 yThrAlaThrLeuGlyGlyThrThrSerPheAlaAsnThrPheThrAsnT 1684
```

2486 ACGGCACTGTGACGCTTTCCGACACGCTTAAGCAAACTAAGCATTC 2535
:::|||||
1684 hclgylalvalthrleu.....Alalysglyserilethrserphe 1697
:::|||||
2536 GCACCTCAAGCGCAATGCTCC.....CTAGCGAATAAAGCAAGTATCCA 2579
||| ||| ::::|
1698 Alalysasnvalthralthrserphevalalasnseralathrilleas 1714
:::|
2580 TTTTGAAGACAGC...CGCTTACCGGACACTCAAGCGGACAGCAAGANA 2626
:::|
1714 npheseranserleualalpheasnserasnilerhrgly..... 1727
2627 CAGCATTAACCTTAAAGACAGCAAGATGAGCCTGCCCTCAGGACGGA 2676
1728glylythrthr 1731
2677 TTAGCAATTTAAACCTTACAAACGCGACCATTAACATTCAGCCCTA 2726
||| ::::|
1732 leu...Thrleuglyalasnlnvalthrtythr..... 1742
2727 TCGCCACGATGCTGACGCGCGCAAAACGCGAGNGTGTAGACGCGCG 2776
1743glythrlyserpetheraspthr... 1750
2777 GCCGCGCTTCGCGCGCTTCCTATTATCCGTACACCGCAACTTCGCTA 2826
1751leuthrleuasnthrthrPheaspglyalala 1761
2827 GAATCCCGTTCAACACGCTGACGGAAGCAAA...TTGAACNGTCA 2873
:::|
1762 lysserglylyasnleuleuileysserglyserthrleuaspleuse 1778
2874 AGGAACATTCCTCGCTTATGTCCGAACCTTCGCTACCGCAAGCGCAAT 2923
:::|
1778 rglvalaserthrleualaleuvalthralthrasnphesapmeta 1795
2924 TGAAGCTGGCGGAAGATTCCGAAGNACTTACACCTTGCGGCTCAACAT 2973
1795 snasnleiserProaspthrlysthrValilleseralaglunhrala 1811
2974 ACCGGCAAGAACCGGTAGCCTGCATCATATTGACGTAAGTGAAGGAA 3023
1812 GlyglyleuuleysProthrserlysglunsnvalysillethrilleasnas 1828
3024 AGACAACAACCGCTGCTCC.....GAAACCTTAAATTCACCGCTGC 3064
1828 naspsnasrghphevalasrphethrPheaspalaserthrleuthrleup 1845
3065 AAAACGACACGCTCGATGCGCGCGCTGACCTTACCAACTATCCGCAAA 3114
1845 healaglunspillealalaaspval.....lle 1854
3115 GACGGGAGATTCCGCTGCATTAATCCGGCAAGAAAGAAAGACTTCGCA 3164
1855 aspglyasrphelalaproglyleuproleu..... 1864
3165 CAAACTCGCAAGCGCAAGCAAGCAAAACAGCGGAGAAAAAGCAACGCGC 3214
1865Alasnilleproasnalaia 1871
3215 AAAGCTTGACCGCTGATTGCGCGCGCGGATGCCGCGAANAAGACA 3264
1871 laasnillelysserleugluleumetglunspalaproasnlyser 1887
3265 GAAAGCGTTGCCGAACGCGCGCGCGNAGCGGGGAAATGTGCGCAT 3314
1888 Asp.....Alaarglinala...Pheasnasnphelyle 1898
3315 TATGCAGCGGAGGAAGAAAAACGCGGTGCGAGGATAAAGACACG 3364
1898 umethrthrProleuglnglu.....Alaaspalathrthr 1910

3365 CNTTGGCGAAACAGCGCGAAGCGGAACCCGCGGTACACCGCCTTC 3414
1910 isoleulle..... 1912
3415 CCGCGCGCGCGCNGCGCGCGGATTTGCCGACCGACGCCCAAC 3464
1913GlnasrpalvalysProser..... 1919
3465 GCAACCTCAACCCCAACCGCAGCGCATGATNAGCGTTATGCAATA 3514
1920Aspthrillealalavalasnasng 1928
3515 GCGGTTGAGTGAATTTCCGCGACGCTCAACAGCGTTTTCGCGTACAG 3564
1928 lnvalalala.....SerasnleiseraserasnillethrAlaleuasn 1942
3565 GACCAATTTGACCGCGCTTTGCC..... 3588
1943 AlaargmetasrplysvalGlnserglyAsnlysglyProvalsersergl 1959
3589GAAGACCGCGCAACGCGNGTTTGACAAAG...NGCATCC 3625
1959 yAspglunspmetasrpalalysrphelyalatrplleserProphervalg 1976
3626 GGNACACCAACACTACCTGCGCAAGATTTCGCGCTTACCGCCACAA 3675
1976 lysasnlatrhnglnlysmeclysnserlleserglytyrlysserasp 1992
3676 ACCGACTCGCGCCCAATCGT.....ATGCAGAAAACTCGG 3713
1993 Thrthrlyglythrilleglypheaspglyphevalserasplenal 2009
3714 CAGCGG.....CGCTGCGGCATCTGTTTCGACA 3745
2009 aleuglyleualatyrthrarglalaaspthraspllelyleuileysna 2026
3746 ACCGACCGGAACANCTTGACGACGCGCATCGGCAACTCGGACGCGCTT 3795
2026 snlysthr.....glyasrplysnasnlyvalgluser 2036
3796 GCCCGCGCGCGCTTTCGGGCAATACGSCATCGGCACTGCACATCG 3845
2037 AsnillethrserleuylrglyleuylrasnvalProtyrlylunsnleuph 2053
3846 CATGACGACGCGCGCGGTGTTACAGCGGCANTTTCAGACGCGATCG 3895
2053 evalglunallealaserlyserasp..... 2062
3896 GAGGCAAAATCCGC.....CGCGCGTCTGCATTACGSCATTACG 3936
2063 ..AsnlysilleargserlysserArgarvalillealathrthrleuglu 2078
3937 ..GCACGATACCGCGCGCTTCGGCGGATTCGGGATCGCAACCGTACAT 3983
2079 ThrvaliletyrlyglnthrAlaasnnglylystlyrlysserllyserlythr 2095
3984 CGGCGCAACGCGCTATTCTGCCAAAAGCGGATACCGCTAC..... 4026
2095 rgllyglnleu.....MetalaaglytyrthrlyMetmetep 2107
4027 ..GAAACGTCAATATCCACCCCGGCTTCGCTTCAACGNTACGN 4074
2107 roglunasnilleasnleu...ThrPro...leualaglyleualarglytyr... 2120
4075 GCGGCGATTAAAGCAGATTATTCATTCAACCGCGCCACACATNTTCAT 4124
2121 Serthrilleys...AsplysglytyrlysgluThrly..... 2132
4125 CACNCTTATTNAGCTG.....TCCATA 4150
2133 ..ThrthrlytyrlyglnasnleuthrvallysglylysnlytyrAsnthrPhea 2149
4151 CCGATGCCGCTTCGGCGCAAGTCCGAACACGCGTCAATAC..... 4191

```

2149 spg1yleuLeuLysValSerSerAsnIleAsnValAsnGluIle 2165
4192 .....GCGTATGGCTCGATTCGGCAAC 4220
2166 VALLEUTHPRGluLeuTyrAlaMetValSptYrAlaPheLysAsnLy 2182
4221 CCGCAGTCGGAATGGCGGTAACCGCAATCAAGTTTCACG 4266
2182 sValSerAla.....IleAspAlaArgLeuGlnIleThr 2194.

seq_name: SwissProt_40:YDEK_ECOLI

seq_documentation_block:
ID YDEK_ECOLI STANDARD; PRT; 1325 AA.
AC P32051; P76140; P77168;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein ydek precursor (ORF).
GN YDEK OR ORF OR B1510.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
CX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takada J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 595-1325 FROM N.A.
RX MEDLINE=94100243; PubMed=8274505;
RA Cartwright P.J., Timms M.W., Lithgow T., Hoef P.B., Hoogenraad N.J.;
RT "An Escherichia coli gene showing a potential ancestral relationship
to the genes for the mitochondrial import site proteins ISP42 and
MOM38.";
RL Biochim. Biophys. Acta 1153:345-347(1993).
CC -1 SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
(Potential).
CC -1 SIMILARITY: TO E. COLI YFAL.
CC -1 SIMILARITY: SOME, TO FUNGAL MITOCHONDRIAL IMPORT SITE PROTEINS
ISP42 AND MOM38.
CC -1 CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
FRAMESHIFT IN POSITION 653.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000248; AAC74583.1; -.

```

```

DR EMBL: D90793; BAA15190.1; ALT_INIT.
DR EMBL: D90794; BAA15197.1; ALT_INIT.
DR EMBL: X73295; CAA51730.1; ALT_FRAME.
DR PIR: S34315; S34315.
DR Ecogene: EC11780; ydek.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Membrane; Lipoprotein; Signal;
KW Complete proteome.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 1 18 HYPOTHETICAL LIPOPROTEIN YDEK.
FT LIPID 19 19 N-ACYL DIGLUCERIDE (POTENTIAL).
FT CONFLICT 884 884 N -> K (IN REF. 3).
FT CONFLICT 1317 1317 M -> S (IN REF. 3).
SQ SEQUENCE 1325 AA; 136514 MW; 26A3A06FA19AD7D CRC64;

```

```

alignment_scores:
Quality: 236.00 Length: 961
Ratio: 0.524 Gaps: 50
Percent Similarity: 46.826 Percent Identity: 20.604

```

alignment_block:

US-09-303-518D-651 x YDEK_ECOLI ..

Align seg 1/1 to: YDEK_ECOLI from: 1 to: 1325

```

634 GATTATTCCTACTCCGCGCATGGTTATTTGGC.....GCCAA 671
77 AspValAlaTyrAsp...AlaTyrLeuValGlyTyrPheGlyThrGlyVal 92
672 TACACATATGCAGGAGTTGGGAAATATATATGCGTATTTGAGCGCGC 721
92 IleAsnIleLeuAlaGlyGlyAsnAlaSerLeuThrIleThrThrs 109
722 ATGTGCGCCATGCCAAGCAGTATGGCCCTATGCCGATGAGGTGGCGCA 771
109 erValIleGlyAlaAsnGlu.....AspSerGluGlyThrVal 121
772 GCGCAGCAGCGGTTCGCGCATGTTATTTATGACAAACAAATTAATAG 821
122 AsnValLeuLeuGlyThrThrPargLeuTyrAspSerGlyAsnAlaAr 138
822 GCTGCTCAACGAGGAGTTTTCACAAACCGGTTACCTTATTCGCGAGGAAA 871
138 gpIroLeuAsn...ValGlyGlnSerGly..... 146
872 ACGGTTTCACGTCGATACGCAAGATTTGTTACGATGACATTTCACA 921
146 ..... 146
922 GCGCATACATACACGCTGCTNTTTGAACCGCGCAGTAAC.....GG 962
147 ....ThreGlyThrLeuAsnIleLysGlnLysGlyHisValAspGly 161
963 ACATTTTTCCTTACATCAACAACAACAGGTACGGGTACGTAACAGAAA 1012
161 YTyLeuAlaArgLeuLysSerSerThrGlyValGlyThrValAsnValG 178
1013 CCAACGAAAGAGTTTCATCAACAAGCTTAAGTACAGACAGTCGACG 1062
178 LngLysGluAsp.....SerValLeuThrThrGlnLeu 188
1063 TTTCAGC.....GATCTTTGAATGAACATGATTAAGA 1094
189 PheGlnIleLeuLysTyrGlyThrGlySerLeuAsnIleThrAspLysG 205
1095 ACCAGTT.....TACGCGCGCAGGAGGTG 1117
205 YTyValThrSerSerIleValAlaIleLeuGlyTyrGlnAlaGly.... 220
1118 TTAATCAGTACCGTCCACAGGTTAAACACAGCGTGAACACCTTTTATC 1167
221 .....SerAsnLysGlnValVal.....Val 227

```

```

1168 GATTACGCGACGCAAACTCATTTATCAAAACATC..... 1206
    :: |||:: ||||| ::|||::
228 GlutylselylglylurthpleuilelysAsnAsnAspSerIlecluph 244
1207 ..... ACCAAGCGCGCGGCGTTGTATTGAAAGTGATTTAA 1246
    |||||::|||::
244 eginileglysnlnglylthrgly..... 252
1247 CGGTCGCGCTGAACAAACGAGGCAAGCGCGGCGTTTCATATC 1296
    :: ||| :::: |||::
253 ..... GluAlaThrIleArgIleuglyglyLeuValThr 263
1297 AGTGAAGACAGTACCGTTACTTGAAAGTAAACGGCGTGCAACGACG 1346
    ::|||::: :::: |||::
264 AlagIuAsnthrlleileglylysnAlaThrGlyle..... 276
1347 CCTGTCCAAATCGCAAGGCGACGCTGCACGTTCAAGCCAAAGGGGAAA 1396
    |||||::|||::
277 ..... GlyThrLeuAsnValGln..... AspG 284
1397 ACCAAGCGCTGATCAGCGTG.....GGCGAC 1422
    ::|||::: |||||::
284 IAspSerValIleThrValArgArgLeuThrAsnGlyTyrPheglyAsn 300
1423 GGTACAGTCATTTTGGATCAGACGAGACGATTAAGGCAAAACACACG 1472
    |||||::: ::|||::
301 GlyThrValAsnIleSerAsnAsnGlyLeuIleAsnAsnlysglu..... 315
1473 CTTTAGTGAATCGCGTGTGTCAGCGGC...AGGGTACGGTGCAACTGA 1519
    ::|||::: |||||::
316 .TyrSerLeuValGlyValGlnIleAspIleSerHisGlyValIleAsnValT 332
1520 ATGCCGTAATCAGTCAAC.....CCCGACAA 1548
    ::|||::: |||||::
332 hraspIysGlyHisThrAsnPheLeuGlyThrGlyGluAlaPheAlaGlyTyr 348
1549 CTGTATTTGGCGTTTCGCGGC...GGACGTTGGATTAAAGCGGCATTC 1595
    ::|||::: |||||::
349 IleTyrIleGlyAspAlaGlyAspGlyGluLeuAsnValSerSerIleGlu 365
1596 GCTTGGTTCACCGTATTCAAAATACGATGGAAGGCGGAGATGTCGNCN 1645
    : ..... LysValAspSerGlyIleIleThrAlaG 375
365 Y..... 399
1646 ATCATATGCCACACACATCCACCGTTACCTACAGGGAATGAAGT 1695
    ::|||::: |||||::
375 lyMetIysGluThrGlyThrGlyAsnIleThrValIlysAspIysAsnSer 391
1696 ATTACACACCGAGTGTGAAGATATACATTAAATTACGACCAAGA 1745
    ::|||::: |||||::
392 ValIleThrAsnLeuGlyThrAsn..... 399
1746 AATTGCTTACACGTTGGTTGGCGAAGAAATACACCAAAACGACAG 1795
    ::|||::: |||||::
400 .LeuGlyTyrAspGly.....HisG 406
1796 GCGCGGTCAACTTGTTTACACCGCGCGACAGAACGACGACCGCGTG 1845
    |||::|||::: |||||::
406 lyGluMetAsnIleSerAsnGlnGlyLeuVal.....Val 417
1846 CTTTCGCGGGAACAATTTAAAC.....GGCAACAT 1877
    ::|||::: |||||::
418 SerAsnGlyIysSerSerLeuGlyTyrGlyGluThrGlyValGlyAsnVal 434
1878 CAGGCAAAACAGCAAACTGTTTACAGCGGACACGACACGCGCAG 1927
    ::|||::: |||||::
434 IserIleThrThrGlyGlyMetIrp.....GluValAsnIysAsnVal 448
1928 CCTACATCATTTAGAACGCGGCGTGTCAAAATAGAGATTCACCAAA 1977
    ::|||::: |||||::
448 alTyrThrThrIleGly.....ValAlaGlyVal..... 457

```

```

1978 GGAGAAATCGTGTGGACACGACTGATCNACGCAAGTTTAAAGCGGA 2027
458 .....G1 458
2028 AAATTTCATATTCAGGGCGGCGGATTTCCGCAATGTGCA 2077
    |||::|||::: |||||::
458 yAsnLeuAsnIleSerAspGlyGlyLysPheValSerIleAsnIleThrP 475
2078 AAGTGAAGCGGATTTGAGCAATCAGCCCAAGTAGTTTGGT 2127
    ::|||::
475 heLeu..GlyAsp..... 478
2128 GTGCGACCGCATCAAGCCATACATCTGACGTTGCGACGTGACNGG 2177
    ::|||::: |||||::
479 .....LysAlaSerGlyIleG1 484
2178 TCTGCAATTTGTGCGAANAANCATTACGAGATTAATGTTCCTT 2227
    |||::|||::: |||||::
484 yThrLeuAsnLeuMetAspAlaThrSerSerPheAspThrValGlyIleA 501
2228 CATTGACTAAGACNGACNTNAGCGCATGTGACGTCGTCGATNACGNT 2277
    ::|||::: |||||::
501 snValGlyAsnPhe..GlySerGlyIleValAsnValSerAsnGlyAla 516
2278 NNTTNAANCTCNCNGGCGTGCNNACTNANAGCAATCTTAGTCAAA 2327
    ::|||::: |||||::
517 ThrLeuAsnSerThrGlyTyrGlyPheIleGlyIlysnAlaSerGlyLy 533
2328 TGCG.....GATPACACGTTAT.....ACAGTCA 2350
    ::|||::: |||||::
533 sGlyIleValAsnIleSerThrAspSerLeuThrAsnLeuLysThrSerS 550
2351 GCCACACGCGC.....ACCCAAACGCGC 2373
    |||::|||::: |||||::
550 erThrAsnAlaGlnLeuLeuGlnValGlyValLeuGlyThrGlyGluLeu 566
2374 AACCTTACCTCGTGGCAATGCCACAGCAACATTATATCAAGCCACATT 2423
    |||||::: |||||::
567 AsnIleThrThrGlyGlyIleValIlysAlaArgAspThrGlnIleAlaE 583
2424 AAC.....GGCAACNCATCGGNTTGGGCAATGCTCA.... 2457
    |||||::: |||||::
583 uAsnAspIysSerIysGlyAspValArgValAspIylGlnSerLeuL 600
2458 .....TTTAATCTAAGCAACAGCGCGCAAAACGCGAGTCTGACG 2499
    |||||::: |||||::
600 euGluThrPheAsnMetIryValGlyThrSerGlyThrGlyThrLeuThr 616
2500 CTTTCGCAACGCTAAGGCAAAAGTA..... 2526
    |||||::: |||||::
617 LeuThrAsnAsnGlyThrLeuAsnValGlnGlyGluValTyrLeuG1 633
2527 .....AGCCAT.... 2532
633 yValPheGluProAlaValGlyThrLeuAsnIleGlyAlaAlaHisGlyG 650
2533 .....TCGCACTCAACGCAATGTCTCCCTAGCGGATTAAGCA..... 2571
    ::|||::: |||||::
650 lualAlaAlaIAspAlaGlyPheIleThrAsnAlaThrIysValGluPhe 666
2572 .....GTATTCATTTTGAANAACAGCGCGCTTAC 2600
    |||||::: |||||::
667 GlyLeuGlyGlnGlyValPheValPheAsnHisThrAsnSerAspAl 683
2601 CGGACAA.....CTCAGCGGCGACGAAGANACAGCATTTAC 2635
    ::|||::: |||||::
683 aglyTyrGlnValAspMetLeuIleThrGlyAspAspIysAspIylYsv 700
2636 ACTTAAAGACAGCAATGACGCTG.....CGTCAAGCAGCAGGAATTA 2679
    ::|||::: |||||::
700 alIleHisAspAlaGlyHisThrValPheAsnAlaGlyAsnThrTyrSer 716
2680 GGCATTTTAACTTGACACGCGCACCATTAACCTCAATTCGCGCTATCG 2729

```



```

ET CONFLICT 985 985 N -> K (IN REF. 1)
ET CONFLICT 1009 1009 L -> S (IN REF. 1)
ET CONFLICT 1013 1013 Y -> S (IN REF. 1)
ET CONFLICT 1182 1182 K -> O (IN REF. 1)
ET CONFLICT 1314 1314 N -> Y (IN REF. 4)
ET CONFLICT 1451 1451 H -> N (IN REF. 1)
ET CONFLICT 1624 1624 E -> D (IN REF. 1)
ET CONFLICT 1628 1628 E -> G (IN REF. 1)
ET CONFLICT 1872 1872 A -> V (IN REF. 1)
ET CONFLICT 1875 1875 T -> P (IN REF. 1)
ET CONFLICT 1878 1879 MS -> LP (IN REF. 1)
ET CONFLICT 1936 1936 E -> A (IN REF. 1)
ET CONFLICT 1965 1970 MTADLP -> ITPLP (IN REF. 1)
ET CONFLICT 1997 1997 G -> R (IN REF. 1)
SQ SEQUENCE 2021 AA: 203328 MW: 327FC42D7CB24668 CRC64;

```

alignment_scores:
 Quality: 234.00 Length: 1401
 Ratio: 0.363 Gaps: 65
 Percent Similarity: 45.967 Percent Identity: 19.914

alignment_block:

US-09-303-518D-651 x OMPA_RICCN ..

Align seg 1/1 to: OMPA_RICCN from: 1 to: 2021

```

622 GACAAACAGCGGATTTACTACTCCGCGCATGTTATATGGCGCA 671
||||| : : : : : : : : : : : : : : : : : : : : :
319 AspsnAlaSerAlaValThrPheThrAsnProValValThrGlyAl 335
672 TACACATATGACGGTGGGAAATATAGCGGTATTAAGTTGACGGCG 721
||||| : : : : : : : : : : : : : : : : : : : : :
335 alleAspsnThrGlyAsnAlaAsnAsnIleValThrPheThrGlyA 352
722 AT.....GTGGCCATGCCAAGCAAGCACTATGCCCTATG 753
|| : : : : : : : : : : : : : : : : : : : : :
352 spSerThrValThrGlyAsnIleGlyAsnThrAsnAlaLeuAlaThrIle 368
754 CGGATTCAGGTGGCGGACGACGCGGTTGCCAATGTTT..... 795
||||| : : : : : : : : : : : : : : : : : : : : :
369 SerValGlyAlaGlyLysAlaThrLeuGlyAlaIleIleLysAlaTh 385
796 .....ATTATGACAAACAAACAAATTAATGCTCTCAAGGAG 835
|| : : : : : : : : : : : : : : : : : : : : :
385 rThrThrLysLeuThrAspAsnAlaSerAlaValThrPheThrAsnPro 402
836 TTTTCAAAACCGGC...TACCTTATTCGGGACGGAACGAGTTTCAG 882
||||| : : : : : : : : : : : : : : : : : : : : :
402 alValValThrGlyAlaIleAspAsnThrGlyAsnAlaAsn..... 415
883 CTGATACGCAAGATTGTTTACGATGACATTTACAGAGCGATACACA 932
415 ..... 415
933 TACCGCTTTTTCGACCGCGCAGTACGACGACATTTTCCTTACATCCA 982
||||| : : : : : : : : : : : : : : : : : : : : :
416 .....AsnGlyIleValThrPheThrGlyA 424
983 ACAACAGCGTACGGGTACGGTACGTAACGAACCAAC..... 1017
424 spSerThrValThrGlyAsnIleGlyAsnThrAsnAlaLeuAlaThrIle 440
1018 .....GAAAAGTNTCAATCCAAAGCTTAAGTACCA 1049
441 SerValGlyAlaGlyLysAlaThrLeuGlyAlaIleIleLysAlaTh 457
1050 GACAGTCCGACTGTTGACGAATCTTGAATGAACATGATTAAGAACCG 1099
||||| : : : : : : : : : : : : : : : : : : : : :
457 rThrThrLysLeuThrAspAsnAlaSerAlaValThrPheThrAsnPro 474
1100 TTTACGGCGGAGGGGTGTTATATGATGATCGTCCAGGTTAAACAGCGT 1149
|| : : : : : : : : : : : : : : : : : : : : :

```

```

474 alValValThrGlyAlaIleAsp.....AsnThrGly 484
1150 GAA...AACCTTCCTTATGATTTAGGACGCAACGCAACTCATCTATC 1196
||| : : : : : : : : : : : : : : : : : : : : :
485 AsnAlaAsnAsnGlyIleValThrPheThrGlyAspSerThrValThrGly 501
1197 AAC.....AACATCAACCAAGCGCGCGG 1222
||| : : : : : : : : : : : : : : : : : : : : :
501 yAsnIleGlyAsnThrAsnAlaLeuAlaThrIleSerValGlyAlaGlyL 518
1223 GTTGTATTTTGAAGTGATTTTACGGCTTCGCCGTGAACCAACGAACG 1272
||| : : : : : : : : : : : : : : : : : : : : :
518 ySalThrThrGlyAlaIleIleLysAlaThrThrThrLysLeuThr 534
1273 TGCGAAGCGGCGGCTTCATATCAGTGAAGACAGTACGCTTCTGGAA 1322
||||| : : : : : : : : : : : : : : : : : : : : :
535 AspsnAlaSerAlaValThrPheThrAsnProValValThrGlyAl 551
1323 AGTAAACGGCGTGCAACGACGCGCTGTCCAAATTCGCAAGGACGCG 1372
||||| : : : : : : : : : : : : : : : : : : : : :
551 alleAspsnThrGlyAsn..... 557
1373 TGCACTTCAAGCCAAAGGGGAAACCAAGCGCTCGATCAGC...GTGGCC 1419
||||| : : : : : : : : : : : : : : : : : : : : :
558 .....AlaAsnAsnGlyIleValThrPheThrGly 567
1420 GACGCTACAGTCATTTTGATCAGCAGCAGACGATTAAGCAAAACAA 1469
||||| : : : : : : : : : : : : : : : : : : : : :
568 AspSerThrVal.....ThrGlyAsnIleGlyAsnThrAs 579
1470 AGCCTTATGTAATCGGCTTGNTCAGCGGCGAGGGGTACGTCACACTGA 1519
||||| : : : : : : : : : : : : : : : : : : : : :
579 nAlaLeuAlaThrIleSerValGlyAlaGlyLysAlaThrLeu..... 593
1520 ATGCCGATTAATCACTTAACCCGACAAACTTATTTGCGCTTCGCGGC 1569
594 .....Gly 594
1570 GGA.....CGTTGATTTAAACGGCGATTC 1595
|| : : : : : : : : : : : : : : : : : : : : :
595 GlyAlaIleIleLysAlaThrThrThrLysLeuThrAspAsnAlaSerAl 611
623 lalIleAspAsnThrGlyAsnAla...AsnAsnGlyIleValThrPheThr 638
1634 CGATGATTCGTCACGATTAATGCGCAACAAACATCCACCGTTACATTA 1683
||||| : : : : : : : : : : : : : : : : : : : : :
623 lalIleAspAsnThrGlyAsnAla...AsnAsnGlyIleValThrPheThr 638
1684 GCGATGAAGATTAATACCAACGAGTGTGAAGATATCAATGACTT... 1731
||||| : : : : : : : : : : : : : : : : : : : : :
639 GlyAsnSerThrValThrGlyAsnIleGly...AsnThrAsnAlaLeuAl 654
1732 .....AATTACGCAAGAAATGCGC...TCAACGCTTGGTTGGCG 1771
||| : : : : : : : : : : : : : : : : : : : : :
654 arThrValAsnValGlyAlaGlyLysAlaThrLeuGlyAlaValIleL 671
1772 AGAAAGTACGACCAAA...ACGAACGGG..... 1797
||| : : : : : : : : : : : : : : : : : : : : :
671 ySalThrThrThrLysLeuThrAsnAlaLeuAlaSerValLeuThrLeuThr 687
1798 CGGCTCAACCTTGTTCACGACCGCGGAGAGACGCGACCCNGCTGCT 1847
||||| : : : : : : : : : : : : : : : : : : : : :
688 AsnValAsnAlaValLeuThrGlyAlaIleAspAsnThrThrGlyValAs 704
1848 TTCGCGGGAACA...AATTAAACGCAACATCAGCAACAAACGCA 1894
||||| : : : : : : : : : : : : : : : : : : : : :
704 pAsnValGlyValLeuAsnLeuAsnGlyAlaLeuSerGlnValThrGlyA 721
1895 AACTGTTTTCACGCGCAGACGACGACCGCCCTACATATCATTTAGCA 1944
||||| : : : : : : : : : : : : : : : : : : : : :
721 snIle.....GlyAsnThrAsnAlaLeuAlaThrIleSerValGly 734

```


1945 ACCGGGTGTCATAAATGGAAGT..... 1968
::: ||| :::: |||
725 AlaGlyLysAlaThrLeuGlyGlyAlaValIleLysAlaThrThrThr 751
1969ATCCAC 1975
751 sLeuThrAspSnaIaSerAlaValThrPheThrAsnProValValValT 768
1976 AAGGAAATGCTGTGGACACGACCTGATCNCGCC..... 2013
||| ||| |||||
768 hrGlyAlaIle.....AspAsnThrGlyAsnAlaAsnAsnGlyIleAla 782
2014 ACGTTAAAGCGAAATTTTC.....CATATTGCGGGGACAGC 2054
||||| ::::: ||| ::::: |||
783 ThrPheThrGlyAspSerThrValThrGlyAsnIleGlyAsnThrAsnAl 799
2055 GGTGATTTCCCGCAATGTGCC.....AAAGTGAAGGC.... 2088
::: ||| ::::: |||
799 AlaValaThrValAsnValGlyAlaGlyLeuLeuThrValGlnGly 816
2089GATGNCATTGACCAATCAGCCCAAGCAAGTTT 2124
||| ||| ::::: |||
816 alValLysSerAsnThrIleAsnLeuThrAspSnaIaSerAlaValThr 832
2125 GGTGCGACCGCATCAAGCCATCTGTACAGCTGC...GACTG 2171
::: ||| ::::: |||
833 PheThrAsnPro.....ValValaThrGlyAlaIleAspAs 845
2172 GACNGGT...CTGACAAATGTGTGGAAMAANCATTACGACGATAAAG 2218
||||| ::::: |||
845 nThrGlyAsnAlaAsnAsnGlyIleValThrPheThrGlyAspSerThrV 862
2219 TGATTGCTTCATGACTAAGACNGACNTNACGGCANTGTNACNTNCC 2268
||| ::::: |||
862 alThrGlyAsnIleGlyAsnThrSnaIaLeuAlaThrIleSerValGly 878
2269 NATNACGNTNNTNAANCNCNGCGNTGNCACNTNANGCAATCT 2318
::: ||| ::::: |||
879AlaGlyLysAlaThrLeuGlyGlyAlaIle 888
2319 TAGTCAATGGCGATACAGCTTATACAGTCCACCAAGCC.... 2361
||| ::::: |||
888 eIleLysAlaThrThrThrLysLeuThrAspSnaIaSerAlaValThrP 905
2361 2361
905 heThrAsnProValValaThrGlyAlaIleAspAsnThrGlyAsnAla 921
2362ACCCAAACGCAACCTTAGCCTGCGGCAATGC 2396
||| ::::: |||
922 AsnAsnGlyIleValThrPheThrGlyAspSerThrValThrGlyAsnIle 938
2397 CCACGCAACATTTAATCAAGCCACATTAAC..... 2427
||| ||| |||||
938 eGlyAsnThrSnaIaLeuAlaThrValAsnValGlyAlaGlyValThrL 955
2428GGACACNCATCGGNTTGGGCATGCTTCATTAACTAAGC 2469
||| ::::: |||
955 euGlnAlaGlySerLeuAspAlaAsnAsnIleAspPheGlyAlaArg 971
2470 AACACGCGCGACAAACGAGTCTGACGCTTCCGCAAC..... 2511
::: ||| |||||
972 SerThrLeuGlnPheAsnGlyProLeuAspGlyGlyAlaAsnAlaIlePr 988
2512GCTAAGCAACGTAAGCCATTCGCGCACTCAAG 2545
||| ||| |||||
988 oTyrtyrPheLysGlyAlaIleAlaAsnGlyAsnSnaIaIleLeuAsnV 1005
2546 GCAATGTCCTAGCGATAGAGCATTCATTCATTT..... 2563
||| ::::: |||
1005 alAsnThrLysLeuLeu.....ThrAlaTyrHisLeuThrIleGlyThr 1019

2584GAAACAGCCGCTTACCGGACAACTCAG 2612
::: ||| ::::: |||
1020 ValAlaGlnIleAsnIleGlyAlaGlyAsnLeuPheAlaIleAspLase 1036
2613 CGCGACAGAAAGACACATTACACTTAAAGACAGCAATG...ACCG 2659
||| ||| |||||
1036 rAlaGlyAspValThrIleLeuAsnAlaGlnAspIleHisPheArgAla 1053
2660 TGGCGTCAGCGACGGAATTAGCAATTA..... 2688
||| ||| |||||
1053 euAspSerAlaLeuValLeuSerAsnLeuThrGlyAlaGlyAlaAsn 1069
2689AACCTGACACGCCACAT 2708
::: ||| ::::: |||
1070 IleLeuLeuAlaIleAspLeuValAlaProGlyValAspGlnGlyThrVa 1086
2709 TACATCAATTCGCGCTATCGC.....CACGATGCTCAG 2743
::: ||| ::::: |||
1086 lValPheAspGlyGlyValAsnGlyLeuAsnIleGlySerAsnValAlaG 1103
2744 GCGCGCAACCGGACAGTGTACAGACGCGCGCGCGCTCGCGCGT 2793
||| ||| ::::: |||
1103 lYalAlaIaArgAsn...IleGlyAspValGlyGlyAsnLysPheAsnThr 1118
2794 TCCCTATTA.....TCCGTTACACCGCAACTCGGTAGAA... 2829
||| ::::: |||
1119 LeuLeuIleTyrAsnAlaValThrIleThrAspAspValAsnLeuIle 1135
2830TCCG 2833
1135 yIleGlnAsnValLeuIleAsnAsnSnaIaAspPheThrSerSerThra 1152
2834 CTTTCAACAGCGTCAGCGTA..... 2853
||| ||| |||||
1152 lPheAsnAlaGlyThrIleGlnIleAsnAspAlaThrTyrThrIleAsp 1168
2854AACGCAATTTGAACNGT...CAAGCAACATTCGCTTTATGTC 2894
||| ||| |||||
1169 AlAsnAsnGlyAsnLeuAsnIleProAlaGlyAsnIleLysPheAla... 1184
2895 GGAATCTTGGCTACCGAAGCGACAAATTTGAAGCTGGCGGAAAGTTCCG 2944
::: |||
1185HisAlaAspAlaGlnLeuIleLeuGlnAsnSerS 1196
2945 AAGSN.....ACTTACACCTGGCGGTACAC.....AAT 2973
||| ||| |||||
1196 eGlyAsnAspArgThrIleThrLeuGlyAlaAsnIleAspProAspAsn 1212
2974 ACCGGCAACGACCGGTAAGCTCGATCAATTAAGCGGTAGTGAAGGAA 3023
::: ||| |||||
1213 AspAspGlnGlyIleValIleLeuAsnSerValThrAla....GlyGly 1227
3024 AGACAAACAAACCGCTGCGAAAACTTAATTTACCCCTGCAAAACGAAC 3073
||| ||| |||||
1227 s.....LysLeuThrIleAlaGlyL 1235
3074 ACGNGATGCGCGGCGGTGCGTTACCAACTCTCGCAAGCGCGAG 3123
||| ||| |||||
1235 ystThrPheGlyGlyAlaHisLysLeuGlnAspIleValPheLysGlyGln 1251
3124TTCGCGCTCATATATCGGTCAA 3146
||| ||| |||||
1252 GlyAspPheGlyThrAlaGlyThrThrPheAsnThrThrAsnIleValIle 1268
3147 AGAACAAAGACCTTTCGACAACTCGGACGAGCAAGCAAAACAGG 3196
::: ||| |||||
1268 uAspIleThrGlyGlnLeuGlnLeuGlyAlaThrThrAlaAsnValVal 1285
3197 CGGAAAGACCAAGCGGCA.....AGC 3219
||| ||| |||||
1285 euPheLysAspAlaValGlnLeuThrGlnThrGlyAsnIleGlyGlyPhe 1301
3220 CTTGACGCGCTGATGCGCGCGCGCATGCCGCAAAAGACAGAAAG 3269

```

1302 LeuAspPheAsnAlaLysAsnGlyThrValThrLeuAsnAsnValAs 1318
13270 CGTTGGCGAAGCGCCGCGCGAGCGGAAATGTCGCATATATAC 3318
1318 nVAlAlaGlyThrValLysAsnThrGlyThrAsnAsnGlyThrLeuI 1335
3319 A.....GGCGAGGAGAGAGAA 3335
1335 LeValLeuGlyAlaSerAsnLeuAsnArgValAlaSerGlyAlaMetLeu 1351
1336 AAACGGGTCAGGC.....GGATTAAGACAGCGC 3367
1352 LysValGlyAlaGlyAsnValThrIleAlaLysGlyGlyAsnValLys 1368
3368 TGGCGAAGACGCGGAGCGGAAAC.....CCGGCGCGGT 3402
1368 eGlyGlyIleGlnGlyThrGlyThrAsnThrLeuThrLeuProAla...H 1384
3403 ACCAGCGCCTCCCGCGCGCGCGCGCGCGCGGATTTGCC..... 3446
1384 IsPheLysLeuThrGlySerIleAsnLysThrGlyGlyAlaLeuLys 1400
3447 .....GCAACGCGAGCCCGCAACCGCAACCTCAAC 3475
1401 LeuAsnPheMetAsnGlySerValSerGlyValAlaGlyThrAlaAl 1417
3476 CCCAGCGAGCGCGACTGATNAGCGGTATAGCCATAGCGGTTGAGT 3525
1417 aaAsnSerValGly.AsplIeThrThrAlaGlyAlaThrSerPheAlaSer 1433
3526 GAATTTCCGCGACGCTCAACAGCGTTTCCGCGTACAGCAATTTGA 3575
1434 SerValAsnAla.LysGlyThrAlaThrLeuGlyGlyThrThrSerPhe 1450
3576 CGCGGTGTTGGCGAGCGCGCGCGCGCGCGT.....GGACAGAGN 3619
1450 IaIstThrPheThrAsnThrGlyAlaValThrLeuAlaLysIleSerIle 1466
3620 GCATCCGCGNACCAACACTACCTGTCGCAATTTCCGCGCTACACCG 3669
1467 ThrSerPheAlaLysAsnValThrAlaThrSerPheValAlaAsnSerAl 1483
3670 C.....ACAAACGACCTGCGCGCAATCGTATGCAGAA 3704
1483 aThrIleAsnPheGlyAsnSerLeuAlaPheAsnSerAsnIleThrGlyS 1500
3705 AA.....ACCTCGAGCGCGCGGTGGCATCTCTTTTGGCACAAC 3748
1500 eGlyThrThrLeuThrLeuGlyAlaAsnGlnValThrThrGlyThr 1516
3749 GGACGCAAAACANCTTCGACGAGCGCATGCGCACTCGCGCAGCGCTTGC 3798
1517 GlySerPheThrAspThrLeuThrIleAsnThrThr..... 1528
3799 CACGCGCGCGTTTGGCGCAATACGCGATCGCAGGTTTCGACATCGCAT 3848
1529 .....PheAspGlyAlaAlaLysSerGlyLysn..... 1538
3849 CAGCAGCGCGCGGTTTACGACGCGCANCTNTCAGACGCGCATCGAG 3898
1539 .....IleLeuIleLysSerGlySerThrLeuAspLeuSer 1550
3899 GCAAAATCCGCGCGCGCGCTGCATTAAGCATTCAGGACGAGCATCCGC 3948
1551 GlyValSerAsnLeuAlaLeuValAlaThrAlaThrAsnPheAsnMetAs 1567
3949 GCGCGTTTCGCGGATTCGCGATCGAACCGTACATCGCGCAGCGCGCT 3997
1567 nAsnIleSerProAspThrLysThrValIleSerAlaGlyThrAla 1583
seq_name: SwissProt_40:PERT_BORPE

```

```

seq_documentation_block:
ID PERT_BORPE STANDARD; PRT; 910 AA.
AC P14283;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Pertactin precursor (Outer membrane protein P.69) (P.93).
GN PRN OR OMP69A.
OS Bordetella pertussis.
OC Bacteria: Proteobacteria: beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-ISOLATE CN2992.
RX MEDLINE=89264462; PubMed=2542937;
RA Charles I.G., Dougan G., Pickard D., Chatfield S., Smith M.,
RA Novotny P., Morrissey P., Fairweather N.F.;
RT "Molecular cloning and characterization of protective outer membrane
RT protein P.69 from Bordetella pertussis.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:3554-3558(1989).
[2]
RP REVISIONS TO 264 AND 332.
RX MEDLINE=92407514; PubMed=1527510;
RA Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;
RT "Cloning, nucleotide sequence and heterologous expression of the
RT protective outer-membrane protein P.68 pertactin from Bordetella
RT bronchiseptica.";
RL J. Gen. Microbiol. 138:1697-1705(1992).
[3]
RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=96196517; PubMed=8609998;
RA Emsley P., Charles I.G., Fairweather N.F., Isaacs N.W.;
RT "Structure of Bordetella pertussis virulence factor P.69 pertactin.";
RL Nature 381:90-92(1996).
CC -1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STAGE OF WHOOPING COUGH.
CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW Mg(2+)
CC CONCENTRATION.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04560; AAA22980.1; ALT_SEQ.
DR PIR: A32560; A32560.
DR InterPro: IPR003992; pertactin.
DR InterPro: IPR003991; pertactin_related.
DR PRINTS: PR01482; PERTACTIN.
DR PRINTS: PR01484; PERTACTINFAMILY.
KW Outer membrane; Signal; Virulence; Repeat.
FT SIGNAL 1 34
FT CHAIN 35 910 P.93.
FT CHAIN 35 711 PERTACTIN (P.69).
FT PROPEP 712 910 POTENTIAL.
FT SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN
FT ADHESION TO VARIOUS EUKARYOTIC CELL
FT LINES).
FT DOMAIN 266 290 5 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
FT REPEAT 266 270 1.
FT REPEAT 271 275 2.
FT REPEAT 276 280 3.
FT REPEAT 281 285 4 (APPROXIMATE).
FT REPEAT 286 290 5 (APPROXIMATE).
FT DOMAIN 579 593 5 X 3 AA TANDEM REPEATS OF P-Q-P.

```



```

460 .....:::
2820 TTGGTAGAATCCGTTTCAACACGCTGACGGTAAACGGCAATGACN 2869
      ::::: |||||
465 agluAa...GlyArgPheLysValLeuThrValAsn...ThrLeuAlaG 480
      |||||
2870 GTCAAGAACATTCGCTTATATGCGAAGCTGTCGCTACCGAAGCGAC 2919
      |||||
480 LysSerLysLeuPheArgMetAsnValPheAlaAspLeuLysSerAsp 496
      |||||
2920 AAATGAGCTGGCGGAAAGTTCCGAGGNACTTACACCTTGGCGGTCA 2969
      |||||
497 LysLeuValValMetGlnAspAlaSerGlyGlnHisArgLeuThrPala 513
      |||||
2970 CAATACCGGCAAGACCCGTAAGCTCGATCATGACGTTAGGAGG 3019
      |||||
513 gAsnSerGlySerLupProAlaSerLAsnThrLeuLeuValGln 529
      |||||
3020 GGAAGACAACACCGCTGTCCGAAACCTTAATTACCTGCAAAAC 3069
      |||||
530 .....ThrProLeuGlySerAlaAlaThrPheThrLeuAlaAsn 542
      |||||
3070 GAA.....CACGTGATCGCGCGCGGTGCGCTTACCAACTCATCGCA 3113
      |||||
543 LysAspGlyLysValAspIleGlyThrTyArgTyArgLeu..... 556
      |||||
3114 AGACGGCGACTTCGCGCTGATATCCGGTCAAAGACAAGAGCTTCCG 3163
      |||||
556 ..... 556
3164 ACAAACTCGGCAAGCAGAGCAAGCAAAAACAGCGGAAAGACAACGCG 3213
      |||||
3214 CAAAGCTTGACGCGCTGATTCGCGCGCGCGATCGCGCAAAAGAC 3263
      |||||
556 ..... 556
3264 AGAAACGTTTGGCGAAGCCGCCGCGCGGAGCGGGAATGTCGGA 3313
      |||||
557 .....AlaAlaAsnGlyAsnGlyGln.TripSerLeu 566
      |||||
3314 TTATGACGCGGAGAGAGAAAGAAAGGGGTGACGCGATTAAGACACG 3363
      |||||
567 Val..... 567
3364 GCATTGGCGAAGACGCGGAGGAAACCGCGCGGNTACACCGCGCTT 3413
      |||||
568 .....GlyAlaLysAlaProAlaProLysProAlaProG 580
      |||||
3414 CCGCGCGCGCGCGCGCGCGGATTTGCGCAACGCGGACGCCAAC 3462
      |||||
580 InProGly.Pro...GlnProPro.GlnProProGlnProGlnProGlu 595
      |||||
3463 ..CCGCAACCTCAACCCCAACCGCACGCGACCTGATNAGCGGTATGCC 3510
      |||||
595 laProAlaProGlnProProAlaGlyArgGluLeu...SerAlaAlaAla 610
      |||||
3511 AAATGCGGTTTGAAGTATTTCCGCGACGCTCAACAGCGTT.....TT 3554
      |||||
611 AsnAlaAlaValAsnThrGlyGlyValGlyLeuAlaSerThrLeuTrp 627
      |||||
3555 CGCGGTACAGAGAGATTTGACCGCGTGTGTCGCAAGACCGCGAAC 3603
      |||||
627 rAlaGlySerAsnAlaLeuSerLysArgLeuGlyGluLeuAlaGlyLeuAsp 644
      |||||
3604 .....GCGGTTTGACAAAGCNCATCCGCAACCAACCAACATCAC 3642
      |||||
644 roAspAlaGlyGlyAlaTrpGlyArgGlyPheAlaGlnArgGlnGlnLeu 660
      |||||
3643 CGTTCCCAAGATTTCCGCGCTACCGCGCAACCAACCGACCTCGCCAA 3692
      |||||

```

```

661 AspaSnArgAlaGlyArgArgPheAspGlnLysValAlaGlyPheGluLe 677
3693 CGGTATGCGAAGAAACCTC.....GGCAGGGGGCGGTGGCA 3730
      |||||
677 uGlyAlaAspHisAlaValAlaValAlaGlyGlyArgTrpHisLeuGlyG 694
      |||||
3731 TCCGTGTTTGCACACACCGACCGCAACCAACNTCTGCGACGAGCGGATCGC 3780
      |||||
694 LysLeuAlaGlyTyThrArgGlyAspArgGlyPheThrGlyAspGlyGly 710
      |||||
3781 AACTGCGCAGCGCTTCCCGACGCGCGCGTTTTCGCGCAATACGCAATCG 3830
      |||||
711 GlyHisThrAspSerValHis.....ValGlyGlyTyAlaThrTy 724
      |||||
3831 CAGGTGACATTCGGCATCCACACGAGCGCGGCTTTAGACGCGGCAATC 3880
      |||||
724 rLeAlaAspSerGlyPheTyLeuAspAlaThrLeuArgAlaSerGly 741
      |||||
3881 TN.....TCAGACGCG.....ATCGAGCGC 3900
      |||||
741 euGluAsnAspPheLysValAlaGlySerAspGlyTyAlaValLysGly 757
      |||||
3901 AAAATCGCGCGCGCGCTGCTGCTATTCAGCGCATTCAGCAGCATACCGCGC 3950
      |||||
758 LysTyArgTrpThr.....HisGlyValGlyLysSerLeuGlnAl 770
      |||||
3951 CGGT.....TTGCGCGGATTCGGCATTCGAACCGTCACTCG 3985
      |||||
770 aglyArgArgPheThrHisAlaAspGlyTyPheLeuGlnProGlnAlaG 787
      |||||
3986 GCGCAACGCGCTATTTCTCCAAAAGCGATTCACCGCTACGAAAC... 4032
      |||||
787 LysLeuAlaValPheArgAlaGlyGlyGlyAlaTyArgAlaAlaAsnGly 803
      |||||
4033 GTCATATTCGCCACCGCGCTTGCCTTCACCGGNTACCGAGCGGCGCAT 4082
      |||||
804 LeuArgValArgAspGlnGlyGlySerValLeuGlyArgLeuGlyLe 820
      |||||
4083 TAAGGAGATTTATTCATTCAAACCGGCGCAACACATTCATTCATCNCCTT 4132
      |||||
820 uGluValGlyLysArgIleGlyLeuAlaGlyGlyArgGlnValGlnPro 837
      |||||
4133 ATTTNAGCGCTGCTCTATACGAGTCCGCTTCGGGCAAGTCCGAACGCG 4182
      |||||
837 yrlLysAlaSerValLeuGlnLupheAspGlyAlaGlyThrValHis 853
      |||||
4183 GTCATACCGCNGTATTTGCTCAGAGATTTGCGCAAAACCGCGACGTGGGA 4232
      |||||
854 ThrAsnGlyIleAlaHisArgThrGlyLeuAlaGlyTyThrArg...AlaG 869
      |||||
4233 ATGGGCGGTAAACCGCAAAATCAAGGTTTCACGCTGTCNCNTCCACGCTG 4282
      |||||
869 uLeuGly.....LeuGlyMetAlaAla 877
      |||||
4283 CCGCGCGCAAGGNCGCCAATCGAAGCG.....CAACACAGCGCGGAGC 4326
      |||||
877 lalauGlyArgGlyHisSerLeuTyAlaSerTyArgIuTySerLysGly 893
      |||||
4327 ATCAATTTAGGCTACCGCTG 4347
      |||||
894 ProLysLeuAlaMetProTrp 900
      |||||
seq_name: SwissProt_40:PERT_BORPA
seq_documentation_block:
ID PERT_BORPA STANDARD: PRT; 922 AA.
AC P24328;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, last sequence update)
DT 01-OCT-1996 (Rel. 34, last annotation update)
DE Pertactin precursor (Outer membrane protein P.70) (P.95).
GN PRN.
OS Bordetella parapertussis.

```

OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 OC Bordetella.
 OX NCBI_Taxid=519;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ISOLATE CN2591;
 RA MEDLINE=91251771; Pubmed=2041476;
 RA Li L.J., Dougan G., Novotny P., Charles I.G.;
 RT "P.70 pertactin, an outer-membrane protein from Bordetella
 paraptensis: cloning, nucleotide sequence and surface expression in
 Escherichia coli."
 RL Mol. Microbiol. 5:409-417(1991).
 CC -1- FUNCTION: AGGLUTININ THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
 CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
 CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
 CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
 CC CONCENTRATIONS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in, no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X54547; CAA38419.1; -;
 DR EMBL; A26124; CAA01786.1; -;
 DR EMBL; A19182; CAA01454.1; -;
 DR PIR; S15204; S15204.
 DR PIR; S14659; S14659.
 DR InterPro: IPR003992; pertactin.
 DR InterPro: IPR003991; pertactin_related.
 DR PRINTS; PRO1482; PERTACTIN.
 DR PRINTS; PRO1484; PERTACTINFAMLY.
 KW Outer membrane; Signal; Virulence; Repeat.
 FT SIGNAL 1 34
 FT CHAIN 35 922
 FT CHAIN 35 647
 FT PROPEP 648 922
 FT SITE 260 262
 FT DOMAIN 266 290
 FT REPEAT 266 270
 FT REPEAT 271 275
 FT REPEAT 276 280
 FT REPEAT 281 285
 FT DOMAIN 575 603
 SO SEQUENCE 922 AA; 95178 MW; 3DF7BC58D4712478 CRC64;
 alignment_scores:
 Quality: 227.00 Length: 1119
 Ratio: 0.473 Gaps: 56
 Percent Similarity: 42.895 Percent Identity: 20.912
 alignment_block:
 US-09-303-518d-651 x PERT_BORPA ..
 Align seg 1/1 to: PERT_BORPA from: 1 to: 922
 1279 GCGCGGGCGTCATATCATGAGAGACGTACCTTGGAAGTAA 1328
 |||||
 58 GYAlaGlyValArgThrAlaThrGlyThrThrIle.....LysValse 72
 1329 GCGCGGGGCAAGACCGCGCTGTCCAAATCGCAAGGACGCTGCACG 1378
 :|||
 72 rGly.....A 74

1379 TTCAGCCAAAGG.....GAAACCCAGGCTCG...ATCAGCTG 1416
 |||||
 74 rGlnAlaGlnGlyValLeuLeuGluAsnProAlaIleuLeuArgPhe 90
 1417 GCGCGGCTACAGTATTTTGATCAGCAGGACGATTAAGCAAAA 1466
 ::|||
 91 GlnAsnGlySerValThrSerSerGlyGlnLeuPheAspGlyValAr 107
 1467 ACAAGCCTTAGGAATTCGGCTTGTCACCGGACG..... 1503
 107 gArgPheLeuGlyThrValThrValLysAlaGlyLysLeuValAlaAspH 124
 1504GCTAGCTG 1512
 124 IsAlaThrLeuAlaAsnValSerAspThrArgAspAspGlyLeuAla 140
 1513 CAAGTATGCCGATATATCATGTTCAAC.....CCGCAACAAGCTTA 1553
 |||
 141 LeuTyValAlaGlyGlnGlnAlaGlnAlaSerIleAlaAspSerThrIle 157
 1554 TTTCGGCTTCGCGCGGCTTGATTAACGGGCAATTCGCTTCGT 1603
 |||||
 157 uGlnGlyAlaGlyValAlaValGlu...ArgGlyAlaAsnValThrV 173
 1604 TCACCGTATTCAAATATCCGATGAAGGGGCGATGATGNCATCAAT 1653
 ::|||
 173 aGlnArgSerThrIleValAspGlyLeuHisIleGlyThr.LeuG 189
 1554 GCCCAACAAACATCCACCGTTACCATTCACGGGAATGAATGTACACA 1703
 :|||
 189 nProLeuGlnProGluAspLeuPro..... 197
 1704 ACCAGTGTGAAGATATATCAATAGACTTAATTACAGCAAGAAATTCCT 1753
 |||
 198Pro 198
 1754 ACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAAGCGGCGCTC 1803
 ::|||
 199 SerArgValValLeu.....GlyAs 205
 1804 AACCTTGTTCACAGCCCGCGGAGAACCCGACCCGCTGC..... 1846
 :|||
 205 pThrSerValThrAlaValAlaProAlaSerGlyAlaProAlaAlaValPheV 222
 1847 ..TTTCGGCGGAGCAATTTAAACGGCAACATCCAGCAACAAACGCGCA 1894
 |||||
 222 aPheGlyAlaAsnGluLeu..... 228
 1895 AACGTGTTTTCAGCGCGAGACCGACACCGCGCTCAATTCATTAGGA 1944
 ::|||
 229ThrValAspGlyGlyHisIleThr..... 236
 1945 AGCGGTTGTCAAATAATGAAAGGTATCC.....C 1973
 |||||
 237GlyGlyArgAlaAlaGlyValAlaAlaMetAspGlyAlaIleValH 252
 1974 ACAAGGAGAAATCGTGTGGGACAAACGACGTGATCCAGCGCTTTAAG 2023
 ||
 252 IsLeuGlnArgAla.ThrIleArgArgGlyAspAlaPro..... 264
 2024 CGGAAATTTCCATATTCAGGGGCGGCGGCGGTGATTTCCGCAATGTT 2073
 |||||
 265AlaGlyGlyAlaValAlaProGlyGlyA 273
 2074 GCCAAAGTGAAGGCGATGNCATTTGAGCAATACGCGCCCAAGCGTTT 2123
 ::|||
 273 IaValProGlyGlyAlaVal.....ProGlyGlyPhe 283
 2124TGGTGTGCGACCGCATCAAGCGCATATCATC 2155
 |||
 284 GlyProLeuLeuAspGlyTTP..... 290
 2156 GTACACGTTCGACTGACNGGTCTGCAAAATGTTGTGCAANAANCATT 2205

```
291 .....TyrGlyValAspValSerasp.SerThrValasp..... 301
2206 ACGGAGATTAAGTGAATGCTTCATGACTAAGACAGACNTNAGCGGCAN 2255
302 .....LeuAlaGlnSerIleValGluAlaProGlnLeuIleVal 314
2256 TGTNAGNCTNNCCNATNAGCNTNNNTNAAANCTCNCNGGCGNTGCNNCAC 2305
314 aala..... 315
2306 TNANAGCAATCTTGAATGGCAATGCGATACAGCTTATACAGTCAAGCCAC 2355
316 .....IleArgAlaGlyArgGlyAlaArgValThrValSerGly 328
2356 AAC.....GCCACCAAAAGCGCAACCTTAGCGCTCGTGGCAATGCCCA 2399
329 GlySerLeuSerAlaProHisGlyAsnValIleGluThrGlyGlyAla 345
2400 AGCAACATTTAAT.....CAAGCCACATTTAAACGGCA 2431
345 aatgargpneProProAlaSerProLeuSerIleThrLeuGlnAlaG 362
2432 ACNCATCGGNTTCGGCAATGCT..... 2454
362 lYAlaArgAlaGlnGlyArgAlaLeuLeuTyArgValLeuProGluPro 378
2455 ...TCATTTAATCTAGCAACAACGCCGCAAAAGCGAGTCTGACGCT 2501
379 ValLysLeuThrLeuAlaGlyAlaGlnGlyAlaGlnGlyAspIleVal 395
2502 TTCGCAACACGCTTAAGCAACGTAAGCCATTCGCACTCAACAGGC... 2547
395 arthrGluLeuProProIle.....ProGlyAlaSerSerGlyProL 409
2548 ..AATGTCCTCCAGCCGATTAAGCGATTTCCATTTTGAACAGCCGC 2595
409 euaspValAlaLeuAlaSerGlnAla.....Arg 418
2596 TTTACCGGACAACCTCAGCGGACAGAGANACAGATTCACCTTAAAGA 2645
419 TrpThrGly.....AlaThrArgAlaValAspSerLeuSerIleAspAs 433
2646 CACGGAATGAGCGCTGCGTCAAGGCAAGCAATTAAGCAATTTAAACCTTG 2695
433 .....AlaThrTrpValMetThrAspAsnSerAsnValGlyAlaLeuArgLeu 449
2696 ACAAGCCCAACATTCACATCAATTCGCTATTCGCGACATGCTGCAGCG 2745
449 ..... 449
2746 GCGCAACCGGAGANGTGTACAGACGCGCGCGCTTCGCGCGCTTC 2795
450 AlaSerAspGlySerVal..... 455
2796 CCTATTATCCGTTACACCGCAACTTCGGTAAATCCGTTTCAACAGCG 2845
456 .....AspPheGlnGlnProAlaGluAla...GlyArgPheLysVal 469
2846 TGACGCTAAGCGCAATTTGAACNGTCAGAAACATTCGCGCTTATATGTC 2895
469 euheValAsp...ThrLeuAlaGlySerGlyLeuPheArgMetAsnVal 484
2896 GAACCTTTCGGCTACCGAAGCAGCAAAATTTGAAGCTGGCGGAAGTTCGA 2945
485 PheAlaAspLeuGlyLeuSerAspLysLeuValValMetArgAspAlaSe 501
2946 AGGNACTTACACCTTGCGGCTCAACAATACCGGCAACGACCGTAAAGCC 2995
501 rGlyGlnHisArgLeuTrpValArgAsnSerSerGlyLeuProAlaSerG 518
2996 TGGATCAATTTGACGTAAGTGAAGGAAAGACACAACCGCTGTCGCA 3045
::: ::: ::::: ||| :::::
518 LysAsnThrMetLeuLeuValGln.....ThrProArgGlySer 530
3046 AACCTTATTTACCCCTGCAAAAGCAA.....CACGTCAATGGCGCGC 3089
531 AlaAlaThrPheThrLeuAlaAsnLysAspGlyLysValAspIleGlyTh 547
3090 GTGGCGTTACCAACCTACCGCAAA...GACGGGAGATTCGCTGCATCA 3136
547 rTyArgTyArgLeuAlaAlaAsnGlyAsnGlyGlnTrpSerLeu.Val 563
3137 ATCCGGTCAAAAGAACAGAGCTTTCGACAAATTCGCGCAAGGCAAGACC 3186
563 ..... 563
3187 AAAAAACAGGGGAAAAAGAACAGCGCAAAAGCTTGACGCGCTGATTGC 3236
563 ..... 563
3237 GGCCGGCGCGATGCCCGGAAAAAGACAGAAAGCGTTGCCGAACGGGCC 3286
564 .....GlyAlaLysAlaProProAlaProLysProAlaPro...GlnPro 577
3287 GCGNCGCAGCGGGGAAATGTCCGATTTATGACAGCGGAGAAAGAGAA 3336
578 GlyProGln..... 580
3337 AAACGGGTGACGCGGATTAAGACAGCGCNTTGGCGAAACAGCGCGAAGC 3386
580 ..... 580
3387 GGAACCCGCGCGGNTACACACCGCTTCCCGCGCGCGCGCGCGCC 3436
581 .....ProGlyProGlnProProGln.ProProGlnProProGlnProPr 595
3437 GGGATTTGCCGCAACG.....CAGCCCGCAA...CGGCAACCTCAA 3474
595 o.GlnProProGlnProProGlnArgGlnProGlnAlaProAlaProGln 611
3475 CCCCACCGCGCGACCTGATNAGCGGTATCCATAGCGGTTGAG 3524
612 ProProAlaGlyArgGlyLeu...SerAlaAlaAlaAsnAlaAlaValAs 627
3525 TGAATTTTCGCGCACGCTCAACAGCGTT.....TTCGCGCTACAGGACG 3568
627 nThrGlyGlyValGlyLeuAlaSerThrLeuTrpTyAlaGlnSerAsnA 644
3569 AATTGACACCGCGTGTTCGCGAAGACCGCGCGCAAC.....GCN 3606
644 lAlaSerLysArgLeuGlyGlnLeuArgLeuAsnProAspAlaGlyGly 660
3607 GTTTGACAAAGCNGCATCCGGMNACCAACACACTACCGTTCGCAAGATT 3656
661 AlaTrpGlyArgGlyPheAlaGlnArgGlnGlnLeuAspAsnArgAlaG 677
3657 CCGCGCCTACCGCCACAAACCGACCTGCGCCAAATCGTATGACAGAAA 3706
677 yArgArgPheAspGlnLysValAlaGlyPheGluLeuGlyAlaAspHisA 694
3707 ACCTC.....GCCAGCGCGCGCGTTCGCGCATCTGTTTGCAC 3744
694 lAlaAlaValAlaGlyAlaGlyArgTrpHisLeuGlyGlyLeuAlaGlyTy 710
3745 AACCGGACCGGAAAAACANTTCGACGAGCGGATCGCAACTCGGACAGGCT 3794
711 ThrArgGlyAspArgGlyPheThrGlyAspGlyGlyLysHisThrAspSe 727
3795 TGCCACAGCGCGCGCTTTTCGGGCATATAGGCAATCGGACGTTGCACATCG 3844
727 rValHis.....ValGlyGlyTyArgAlaThrTyIleAlaAsnSerG 741
3845 GCATCAGCAGCGCGCGGCTTTTACGAGCGCGCANTCN..... 3882
741 LysPheTyLeuAspAlaThrLeuArgAlaSerArgLeuGlnAsnAspPhe 757
```

```

3883 .....TCAGACGC.....ATCGAGGCAAAATCCGCGCG 3914
758 LysValAlaIaIySerArgIlyTrAlaValAlaIySGIlyuTrAlaIyTrH 774
3915 CGTGGTCGATTCAGGCGATTCAGGCGATACCGC...GCCGGTTTGGCG 3961
774 scIyValGIyValSerLeuGIuAlaIyArGArGrPheAlaHAlaIAspG 791
3962 GATTTCGGCATCGAAGCGTACATACGGCGCAACCGCATTTTCGTCGAAAA 4011
791 LyrTrPheLeuGIuTrOGIAlaIaGIuLeuAlaValAlaPheArGValGIyGI 807
4012 GCGGATTACCGCTACGAAAC...GTCAATATCGCCACCGCGCTGTGG 4058
808 GIyAlaTrArGAlaIaAsnGIyLeuArGValAlaAspGIuGIyLysE 824
4059 GTTCACACCGTACCGGCGGCGATTAAGCGAGTATTCATTCGAAACGG 4108
824 rSerValIleuGIyArGIuLeuGIyLeuGIyValGIyLysArGIleGIuLeu 841
4109 CGCAACACATTCATCCATCCATTCATTCATTCATTCATTCATTCATTCAT 4158
841 IaGIyGIyArGIuValGIuGIuGIuGIuGIuGIuGIuGIuGIuGIuGIu 857
4159 GCTTCGGGCAAGATCCGCAACCGCGTCAATACCGCGTATTCGCTCAGGA 4208
858 PheAspGIyAlaGIyTrAlaIyTrAlaIyTrAlaIyTrAlaIyTrAlaIy 874
4209 TTTCGGCAAAACCGCGATTCGCGATTCGCGATTCGCGATTCGCGATTC 4258
874 uLeuArGIyGIyTrArG...AlaGIuLeuGIy..... 883
4259 GTTTCACGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4308
884 .....LeuGIyMetAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 897
4309 GCG.....CAACACAGCGCGGCGCATCAATTACGCTACCGCTGG 4347
898 AlaSerTrGIuTrGIuTrGIuTrGIuTrGIuTrGIuTrGIuTrGIuTrGIu 912
seq_name: SwissProt_40:AG43_ECOLI

seq_documentation_block:
ID AG43_ECOLI STANDARD; PRT; 1039 AA.
AC P39180; P75614; P97241; Q46771.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigen 43 precursor (AG43) (Flufling protein).
GN FlU OR B2000.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
NCBI_TaxID=562;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655.
RX MEDLINE=97426617; PubMed=9278503.
RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=K12.
RX MEDLINE=97251358; PubMed=9097040.
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakase S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,

```

```

RA Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=ML 308-225.
RA Henderson I.R., Owen P.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
[4]
RN PRELIMINARY SEQUENCE OF 53-78.
RP STRAIN=ML 308-225.
RX MEDLINE=97291704; PubMed=2661530.
RA Catfrey P., Owen P.;
RT "Purification and N-terminal sequence of the alpha subunit of antigen
RT 43, a unique protein complex associated with the outer membrane of
RT Escherichia coli.";
RL J. Bacteriol. 171:3634-3640(1989).
[5]
RN SEQUENCE OF 53-63.
RP STRAIN=K12 / EMG2.
RX MEDLINE=97443975; PubMed=9298646.
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313(1997).
[6]
RN GENE NAME.
RP MEDLINE=97257509; PubMed=9103983.
RA Henderson I.R., Meehan M., Owen P.;
RT "Antigen 43, a phase-variable bipartite outer membrane protein,
RT determines colony morphology and autoaggregation in Escherichia coli
RT K-12.";
RL FEMS Microbiol. Lett. 149:115-120(1997).
CC -1- FUNCTION: CONTROLS COLONY FORM VARIATION AND AUTOAGGREGATION. MAY
CC -1- FUNCTION: AS AN ADHESIN.
CC -1- SUBUNIT: HETERODIMER OF THE PERIPHERAL MEMBRANE PROTEIN (ALPHA
CC CHAIN) ANCHORED TO THE INTEGRAL OUTER MEMBRANE PROTEIN (BETA
CC CHAIN).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE ASSOCIATED.
CC -1- SIMILARITY: TO ADHESIN AIDA-I AND TO BORDETELLA PERTACTIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC
DR EMBL: A8000291; AAC75061.1; ALT_INIT.
DR EMBL: D90838; BAA15825.1; ALT_INIT.
DR EMBL: D90839; BAA15832.1; ALT_INIT.
DR EMBL: U24429; AAB47869.1; -.
DR Ecocore: EG12686; flu.
KW Outer membrane; Signal; Complete proteome.
FT SIGNAL 1 52
FT CHAIN 53 551
FT CHAIN 552 1039
FT VARIANT 2 2
FT VARIANT 41 42
FT VARIANT 46 46
FT VARIANT 157 157
FT VARIANT 188 188
FT VARIANT 303 305
FT VARIANT 320 320
FT VARIANT 372 372
FT VARIANT 493 493
FT VARIANT 497 497
FT VARIANT 585 585
FT VARIANT 709 709
FT VARIANT 721 721
FT VARIANT 721 721

```



```

3927 CGGATTCAGGCGACGATACCGCGCGCTTCGGCGGATTCGCATCGAAC 3976
      : : : : :
957 pTPrValGlnPro..... 961
3977 CGTACATCCGCGCAACGGCATTTCTCCAAAAAGCGATTAACCGCTAC 4026
      : : : : :
962 .....SerValIleArgThrPheSerSerArgGlyAspMetArg... 974
4027 GMAACGTCAATATCGCCACCGCGCTTTCGCTTCAACGAGTACCGNGC 4076
      : : : : :
975 .....ValGlyThrSerThrAlaGly..... 981
4077 GGGCATTAAGCAGATATTCATTCACAAACGGCGCAACACATNCCATCA 4126
      : : : : :
982 .....SerGlyMetThrPheSerProSerGlnAspGlyThr.... 993
4127 CNCCTATTATTAAGCCTGTCTATACGATCCGCTTCGGCGCAAGTCGA 4176
      : : : : :
994 .....SerLeuAspLeuGlnAlaGlyLeuGlnAlaArgValArg 1006
4177 ACAGCGCTCAATACCGCGCTATTCGCTCAGGATTCGCGCAAAACCGCAG 4226
      : : : : :
1007 GlnAsnIleThrLeuGlyValGlnAla..... 1015
4227 TCGGGAATGGGGCGTAACGCCGAATCAAGTTTCACGCTGTCNCC 4276
      : : : : :
1016 .....GlyTyrAlaIleHisSerValS 1022

```

seq_name: SwissProt_40:PERT_BORBR

seq_documentation_block:

ID PERT_BORBR STANDARD; PRT; 911 AA.

AC 003035;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Pertactin precursor (outer membrane protein P.68) (P.94).

PRN.

OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).

OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;

OX NCBI_Taxid=518;

RN [1]

RP SEQUENCE FROM N.A. AND SEQUENCE OF 35-44.

RC STRAIN-ISOLATE CN7531;

RL MEDLINE=92407514; PubMed=1527510;

RA Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.;

RT "Cloning, nucleotide sequence and heterologous expression of the

protective outer-membrane protein P.68 pertactin from Bordetella

bronchiseptica.";

RT J. Gen. Microbiol. 138:1697-1705(1992).

CC -1- FUNCTION: AGGLUTININ THAT BINDS TO EUKARYOTIC CELLS; A PROCESS

MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN

BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.

CC -1- SUBUNIT: MONOMER.

CC -1- SUBCELLULAR LOCATION: Outer membrane.

CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.

CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW Mg(2+)

CC CONCENTRATIONS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation-

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC EMBL; X54815; CAA38584.1; -.

```

DR EMBL: A19180; CAA01453.1; -
DR PIR: A47675; A47675.
DR InterPro: IPR003992; pertactin.
DR InterPro: IPR003991; pertactin_related.
DR PRINTS: PRO1482; PERTACTIN.
DR PRINTS: PRO1484; PERTACTINFAMILY.
KW Outer membrane; signal; virulence; Repeat.
FT SIGNAL 1 34
FT CHAIN 35 911 P.94.
FT CHAIN 35 712 PERTACTIN (P.68).
FT PROPEP 713 911 POTENTIAL.
FT SITE 260 262 CELL ATTACHMENT SITE (INVOLVED IN
ADHESION TO VARIOUS EUKARYOTIC CELL
LINES).
FT SITE 701 703 CELL ATTACHMENT SITE (POTENTIAL).
FT DOMAIN 266 275 3 x 5 AA TANDM REPEATS OF G-G-A-V-P.
FT REPEAT 266 270 1.
FT REPEAT 271 275 2.
FT REPEAT 276 280 3 (APPROXIMATE).
FT REPEAT 570 601 7 x 3 AA REPEATS OF P-Q-P.
SQ SEQUENCE 911 AA; 93995 MM; 3078DF6C2D987A1 CRC64;

```

alignment_scores: Quality: 224.00 Length: 1110
Ratio: 0.475 Gaps: 51
Percent Similarity: 42.523 Percent Identity: 20.000

alignment_block:

US-09-303-518d-651 x PERT_BORBR

Align seg 1/1 to: PERT_BORBR from: 1 to: 911

```

1279 GGGCGGGGGTTCATATCAATGTAAGACAGTACGTTACTTGGAAATGTA 1328
      |||||
58 GlyAlaGlyAlaArgThrAlaThrGlyThrIle.....LysValSer 72
1329 CGGCGTGGCAAGACCGCGCTGTCAAATCGGCAAAAGCAGCGTCCAGC 1378
      : : : : :
72 rGly.....A 74
1379 TTCAGCCAAAGG.....GAAACCAAGCTCG...ATCAGCGTG 1416
      : : : : :
74 rGlnAlaGlnGlyValLeuLeuGlnAsnProIleAlaGluLeuArgPhe 90
1417 GCGAGCGTACAGTATTTGATCAGCAGCGAGCATTAAGCAAAAA 1466
      : : : : :
91 GlnAsnGlySerValThrSerSerGlyGlnLeuPheAspGlyValArg 107
1467 ACAAGCCTTAGTGAAATCGGCTTGNTCAGCGCAGGGGTACGGTCAC 1516
      : : : : :
107 gArgPheLeu.....GlyThrValThrV 115
1517 TGAATGGCGAATATCACTTCAACCCGCAAAACTATTCGCGCTTCGC 1566
      : : : : :
115 allysAla..... 117
1567 GGGCGAGCTTGGATTAACGGGCGATTCGTTGCTCCAGCGTATCA 1616
      : : : : :
118 ...GlyLysLeuValAlaAspHisAlaThrLeuAlaAsnValSerAspTr 133
1617 AAATACCGATGAAGGGCGATGTCGATGTCATTAATGCAACAAACAT 1666
      : : : : :
133 rArgAspAspAspGlyIleAlaLeuTyrValAlaGlyGluGlnAlaGln 150
1667 CCACCGTT.....ACCATTAACGAGATGAATGATTAACAACCG 1707
      : : : : :
150 lAsnIleAlaAspSerThrLeuGlnGlyAlaGlyValArgValGlu 166
1708 AGTGTAAAGATATCAATAGACTTAATTAACGAAAGAAATTCCTACAA 1757
      : : : : :
167 ArgGlyAlaAsnValThr..... 172

```

```
1758 CGGTGGTTGGCGAGAAAGATACGACCAAAACGACGGCGCTCAAC 1807
      ::::: ||| ::::: ||| ||:::
173 .....ValGlnArgSerThrIleValAspGlyGlyLeuHis 185
      ::::: ||| ::::: ||| ||:::
1808 TT.....GTTACCAAGCCCGCGGACGAGAACCGACCCCGCTG 1845
      ::::: ||| ::::: ||| ||:::
185 IeGlyThrLeuGlnProLeuGlnProGluAspLeuProProSerArgVal 201
      ::::: ||| ::::: ||| ||:::
1846 CTTTCGGCGGAGCAAAATTTAAAGCGCAACATCAGCCAAACAAACGGCA 1895
      ::::: ||| ::::: ||| ||:::
202 ValLeuGlyAspThrSerValThrAla.....ValProIleSerGly.. 215
      ::::: ||| ::::: ||| ||:::
1896 ACTGTTTTCAGCGCGGACGACCGACCGCCCTACATCATTTAGCA 1945
      ::::: ||| ::::: ||| ||:::
216 .....AlaProIleAlaIleValSerValIleGlyAla 225
      ::::: ||| ::::: ||| ||:::
1946 GCGGGTGTCAAAATGCAAGGTATCCCAAGAGAAATCGTGTGGAC 1995
      ::::: ||| ::::: ||| ||:::
225 IaAsnGluLeuThrValAspGly..... 232
      ::::: ||| ::::: ||| ||:::
1996 AACGACTGATCMACCGACGTTTAAAGCGAAATTCATATTCAAGG 2045
      ::::: ||| ::::: ||| ||:::
233 .....GlyHisIleThrGln 237
      ::::: ||| ::::: ||| ||:::
2046 CGGCGAGCGGTGATTTCCCGCATGTGCCAAAGTGGAGCGATGNC 2095
      ::::: ||| ::::: ||| ||:::
237 GcIlyrGlnAlaIle.....GlyValAla..... 244
      ::::: ||| ::::: ||| ||:::
2096 ATTTAGCAATCAGCCCGACAGCTTTTGGTGTGCGACCCGATCAAGC 2145
      ::::: ||| ::::: ||| ||:::
245 .....AlaMetAspGlyAlaIleValHisLeuGln 254
      ::::: ||| ::::: ||| ||:::
2146 CATACATCTGTACAGCTTCGGAC..... 2169
      ::::: ||| ::::: ||| ||:::
255 ArgAlaThrHisIleArgArgGlyAspAlaProIleGlyAlaValProGln 271
      ::::: ||| ::::: ||| ||:::
2170 .....TGACNGGTC 2179
      ::::: ||| ::::: ||| ||:::
271 GcIlyAlaValProGlyGlyPheGlyProLeuIleAspGlyTrpArgGlyVal 288
      ::::: ||| ::::: ||| ||:::
2180 TGACAAATTTGTGCAANAANCAATTACCGACGATAAAGTGTGCTTCA 2229
      ::::: ||| ::::: ||| ||:::
288 al.....AspValSerAspSerThrVal..... 295
      ::::: ||| ::::: ||| ||:::
2230 TTGACTAGACGACGACNTNAGCGGCGANTGTNAGCTNNCCNATNAGCNTNN 2279
      ::::: ||| ::::: ||| ||:::
296 .....AspLeuAla 298
      ::::: ||| ::::: ||| ||:::
2280 TTNAANCTCNCNGGCGTGCNMCATNANGCAATCTTAGTGCAAATG 2329
      ::::: ||| ::::: ||| ||:::
298 agInserIleValGlnAlaProGlnLeuGlyAlaIleIleArgAlaGlyAla 315
      ::::: ||| ::::: ||| ||:::
2330 GCGATACAGCTTATACAGTCCGACCAAC.....GCCACCCAAACGGC 2373
      ::::: ||| ::::: ||| ||:::
315 rgGlyAlaIleArgValThrValSerGlyGlySerLeuSerAlaProHisGly 331
      ::::: ||| ::::: ||| ||:::
2374 AACCTTAGCCTCTGGGCAATGCCCAACCAACATTTAAT..... 2412
      ::::: ||| ::::: ||| ||:::
332 AsnValIleGluThrGlyGlyAlaIleArgArgPheProProProIleAla 348
      ::::: ||| ::::: ||| ||:::
2413 .....CAAGCCACATTTAAAGCGCAACNATCGGANTTGGGCAATGCT. 2454
      ::::: ||| ::::: ||| ||:::
348 rProLeuSerIleThrLeuGlnAlaGlyAlaIleArgAlaGlnIleArgAla 365
      ::::: ||| ::::: ||| ||:::
2455 .....TCATTTAATCTAAGCAACGAC 2475
      ::::: ||| ::::: ||| ||:::
365 euLeuTyrArgValLeuProGluProValIlySerLeuThrLeuAlaGlyIly 381
      ::::: ||| ::::: ||| ||:::
2476 GCCGCAAAAAGCGAGTCTGACGCTTTCGACAAAGCTAAGCAAAAGCT 2525
      ::::: ||| ::::: ||| ||:::
382 AlaGlnGlyGlnGlyAspIleValAlaIleThrGluLeuProProIle..... 396
      ::::: ||| ::::: ||| ||:::
2526 AAGCATTCGCGACCTCAACGGC.....AATGTCTCCCTAGCGATTAAG 2569
      ::::: ||| ::::: ||| ||:::
      ::::: ||| ::::: ||| ||:::
397 ....ProGlyAlaSerSerGlyProLeuAspValAlaLeuAlaSerGlnAla 412
      ::::: ||| ::::: ||| ||:::
2570 CAGTATTCATTTTGAACACGCCGCTTATACCGGACACTCAGCGGCAC 2619
      ::::: ||| ::::: ||| ||:::
412 Ia.....ArgThrProGly.....AlaThrArg 419
      ::::: ||| ::::: ||| ||:::
2620 AAGGANAACGATTTACACTTAAAGACGGAATGACGCTGCCGACG 2669
      ::::: ||| ::::: ||| ||:::
420 AlaValAspSerLeuSerIleAspAsnAlaThrTrpValMetThrAspAs 436
      ::::: ||| ::::: ||| ||:::
2670 CAGCAATTCAGCAATTTAAACCTTGACACGCGACCATTAACCTCAAT 2719
      ::::: ||| ::::: ||| ||:::
436 nSerAsnValGlyAlaLeuArgLeu..... 444
      ::::: ||| ::::: ||| ||:::
2720 CCGCCTATCGCCACGATGTGACAGCGCGCAACCGGACAGTGTACAG 2769
      ::::: ||| ::::: ||| ||:::
445 .....AlaSerAspGlySerVal..... 450
      ::::: ||| ::::: ||| ||:::
2770 ACGCCGCGCGCGCTTCCGCGCTTCCTTATTCGGTTACACCGGCAC 2819
      ::::: ||| ::::: ||| ||:::
451 .....AspPheGlnGlnProAla 456
      ::::: ||| ::::: ||| ||:::
2820 TTCGTAGAAATCCGTTTCAACACGCTGACGCTAAACGGCAATGTAACN 2869
      ::::: ||| ::::: ||| ||:::
456 agIleAla..GlyArgPheIlySerIleuMetValasp..ThrLeuAlaG 471
      ::::: ||| ::::: ||| ||:::
2870 GTCAGGACATTCGCTTTATGTGCGAACTCTTCGCGCTACGACGAC 2919
      ::::: ||| ::::: ||| ||:::
471 IySerGlyLeuPheArgMetAsnValPheAlaAspLeuGlyLeuSerAsp 487
      ::::: ||| ::::: ||| ||:::
2920 AAATGGAAGCTGGGGAAGTCCGAAGGACCTTACACCTTGCGGCTCA 2969
      ::::: ||| ::::: ||| ||:::
488 IyLeuValValMetAlaGaspAlaSerGlyGlnHisArgLeuValAla 504
      ::::: ||| ::::: ||| ||:::
2970 CAATCCGCGGACGAAACCGCTAAGCCTCGATCAATACGCTAGTGAAG 3019
      ::::: ||| ::::: ||| ||:::
504 gAsnSerGlySerGlnProIleSerGlyAsnThrMetLeuValGln.. 520
      ::::: ||| ::::: ||| ||:::
3020 GGAAGACAAACACCGCTCCGAAACCTTAATTCACCTCGCAAAAC 3069
      ::::: ||| ::::: ||| ||:::
521 .....ThrProArgIlySerAlaIleThrPheThrLeuAlaAsn 533
      ::::: ||| ::::: ||| ||:::
3070 GAA.....CAGCTGATGCGGCGGCTGTACCACTCATCCGCA 3113
      ::::: ||| ::::: ||| ||:::
534 IyAspGlyIyValAspIleGlyThrTyrArgTyrArgLeuAlaIleAs 550
      ::::: ||| ::::: ||| ||:::
3114 A..GACGGGAGTTCGCGCTGCATATCCGTCAAAAGAACAGACTTT 3160
      ::::: ||| ::::: ||| ||:::
550 ngIyAsnGlyGlnTrpSerLeuVal..... 558
      ::::: ||| ::::: ||| ||:::
3161 CCGACAAACTCGGCAAGGACGAAAGCCAAAAAACAGCGGAAAAAGACAC 3210
      ::::: ||| ::::: ||| ||:::
558 ..... 558
      ::::: ||| ::::: ||| ||:::
3211 GCGCAAGCCTTGACGGGTGATTTGGCGCGGCGGATGCCGCAAAA 3260
      ::::: ||| ::::: ||| ||:::
559 .....GlyAlaIlySalProProA 565
      ::::: ||| ::::: ||| ||:::
3261 GACGAAAGCGTTGCCGACCGCGCGCGGACGCGGAAATGTGCG 3310
      ::::: ||| ::::: ||| ||:::
565 IapTrpIySalProIlePro...GlnProGlyProGlnProGly..... 577
      ::::: ||| ::::: ||| ||:::
3311 GCATTTATGCAAGCGGAGAGAAAGAAAAACGGGTGCAGCGGATAAGAC 3360
      ::::: ||| ::::: ||| ||:::
577 ..... 577
      ::::: ||| ::::: ||| ||:::
3361 AGCGCNTTGGCAAAAGACGCGAAGCGAAACCCGCGGNTACACCGC 3410
      ::::: ||| ::::: ||| ||:::
578 .....Prog 579
      ::::: ||| ::::: ||| ||:::
3411 CTTTCCCGCGCGCGCGCGCGCGGATTTTGCAGCAACGCGACGCC 3460
      ::::: ||| ::::: ||| ||:::
```


576GlnGlyAlaPheValLysThrThrIng 585
1304 ACAGTACCGTTACTTGGAAAGTAAAGCGCGTGGCAAAAGACCGCTGTC 1353
.....
585 LysPValVal.....IleAspAlaLeuSerGluThrIleSer 598
1354 AAAATCGGCAAA.....GGCAGCGTGCAGCTTCAGCCAAAGGGGAAA 1397
|||||
599 LysIleAspGluArgThrGlyThrAlaPheAsnIleThrLysSerSerH 615
1398 CCA.....GGTCGATCAGCGTGGCG 1420
.....
615 sLysAsnGluThrAsnLysGlnThrSerThrGlySerGluLeuIleSerA 632
1421 ACGGT.....ACAGTCATTTGGATCGAGCGAGACGATAAAGGCAA 1464
|||
632 spAlaGlnLeuThrValValSerGlyAsnAspValAsnValIleGlySer 648
1465 AAA.....CAAGCCTTAGTGAATCGCTTGNTCAGCGGAGGAGTACGT 1511
.....
649 LeuIleLysSerAlaAspLysLeuGlyIleHsSer..LeuGlyAspI 664
1512 GCACTGAATGCCGATTAATCAG.....TTCAACCCGACAAACTCT 1552
|||
664 eAsnValLysSerAlaGlnGlnValThrLysIleAspAspGluThrS 681
1553 ATTGGCGCTTCGGCGGAGCTTGGATTAAAGCG.....CATTCG 1596
.....
681 eLeuAlaIleThrGlyHisAlaLysGluValGluAspLysGlnThr 697
1597 CTTTCGTTCCACCGTATTCAAAATACGATGAGGCGGATGATGNCNA 1646
|||
698 AlaGlyPheHsIleThrHisThrAsnLys..... 708
709AsnThrSerThrGluThrGluGlnAlaAsnSerThrI 721
1697 TTACACACCGAGTGTAAAGATATCAATACATTAATTAACAGCAAGAA 1746
|||
721 Le.....SerGlyAlaAsnValAsp..LeuGlnAlaAsnLysAsp 733
1747 ATTGCCTACACGGTTGGTTGGCGAGAAAGATACGACCAAAACAGACGG 1796
|||
734 ValThrPheAlaGly.....SerAspLeuLysThrThrAlaGlyAsnAl 748
1797 GGGGCTC.....AACCTGTTTAC.....CAGCCG 1822
|||
748 AserIleThrGlyAspAsnValAlaPheValSerThrGluAsnLysLysG 765
1823 CGCAGAAAGACCGCACCCNGCTGCTTCGCGGGAACAATTAACAGGC 1872
|||
765 lntHrAspAsnThrAspThrThrIleSerGlyPheSerThrLys 781
1873 AACATACCGCAAAACAGCGCAACTGTTTTCACGCGGACGACGACAC 1922
|||
782 GlyValAspLysValGlySerLysAlaAspPheGln..... 793
1923 GCACGCGCTCAATCATATTAGAAAGCGGCTGTCAAAAATGGAAGTATCC 1972
|||
794 .TyrAspLysGlnHisThrGlnThrGluValThrLysAsnArgGly...S 809
1973 CACAAAGAGAAATCGTGTGGACAAACGACTGATACCGCACGTTTAA 2022
|||
809 eGlnThrGluValAlaGlyAsp.....LeuThrIleThrAlaAsn 822
2023 GCGGAAATTTCCATATTCAGGGCGGAGGCGGATTTCCGCAATCT 2072
|||
823 LysAspLeuLeuHisGluValAlaSerHisHisValGluGlyArgGly 839
2073 TGCCAAAGTGAAGCGATTTGNCATTTGAGCAATCAGCCCAAGAGTTT 2122
|||
839 nGluSerGlyGluAsnIleGlnHisLeu..... 848
2123 TTGGTGTGCGACCGCATCAAGCCATACATCTGTACAGCTTCG..... 2166
|||
849AlaValAsnAspSerGluThrSerLysThrAspSerLeuAsn 862
2167GACTGACNGCTGTGACAAATGTG 2189
|||
863 ValGlyIleAspValGlyValAsnLeuAspGlyThrValThrAlaArgAspAlaIle 879
2190 TTGTGAAANAANACATTACGAGCATTAAGTATGCTTGTATGACAAACA 2239
|||
879 oValLysLysAlaIleGluAspGly.....ValAsnThrThrLysP 893
2240 CNGACNTNACGGCAGTGTNAGCTNCCNATNACGNTNNTTNAAMCTC 2289
|||
893 roGlyAsnAsnThrAspLeuThrLysLysValThrAlaArgAspAlaIle 909
2290 NCNGGCGCTGCNNCCTMAAN..... 2310
|||
910 AlaAsnLeuAlaAsnLeuSerAlaLeuGluThrProAsnValGlyValG 926
2311GGCAATCTTAGTGCAAATGCGCATACAGT. 2340
|||
926 uValGlyIleLysGlyGlyLysGlnGlnSerGlnThrAspSerGlnA 943
2341 ..TATACAGTCACGCAACAGCC..... 2361
|||
943 lValSerThrSerIleAsnAlaGlyLysIleAspIleAspSerAsnAsn 959
2362ACCCAAACGGCAACCT 2378
|||
960 LysLeuHisAspGlnGlyThrHisThrGlnSerThrGlnGlnGly... 975
2379 TAGCCTGTGCGCAATGCC.....C 2398
|||
975 eSerLeuThrAlaAsnThrHisThrSerGluAlaThrLeuAspLysHisG 992
2399 AACCAACATTTAATCAAGCCATTAACGCGCAACMCATCGCTTGGGCG 2448
|||
992 lntHrThrPheHisGluThrLysGlyGlyGlyGlnIleGlyValSerThr 1008
2449 AATGCTCATTTAATCTAAGCAACACCGCGCAAAACGCGAGTGTGAC 2498
|||
1009 LysThrGlySerAspIleThr.....ValAlaIleLysGlyGlnGly 1023
2499 GCTTTCGACACGCT.....AAGCAAAAGTAAAGCATTCG 2536
|||
1023 nThrThrAspAsnAlaLeuMetGluThrLysAlaLysGlySerGlnPheT 1040
2537 CACTCAACGGCAATGCTCCCTAGCGCATTAAGGAGATTCATTTGGA 2586
|||
1040 hrSerAsnGlyAspIleSerIleAsnValGlyGluAsnAlaHisThrGlu 1056
2587 AACAGCCGCTTACCGGACAACCTCAGCGGACGAGCAAGANACACATTA 2636
|||
1057 GlyAlaGlnPheAspAlaGln..... 1063
2637 CTTAAAGACAGCGAATGACGCTGCGCTCAGCGACGGAATTAGCGAATT 2686
|||
1064LysGlyLysThrValIleAsnAlaGly.....GlyAspI 1075
2687 TAAACCTTGACCAAGCCACCATTTACACTCAATTCGCGTATCGGACGAT 2736
|||
1075 eutHrLeuAlaGlnAlaThr..... 1081
2737 GCTCAGGCGCGCAACCGGAGNGTCTCAGACACGCGCGCGCTTC 2786
|||
1081 1081
2787 GCGCGCTTCCTATTATCCGTTACACGCCCACTTCGTTGAATCCGCTT 2836
|||
1082AspThrHisSerGluSerGlnS 1089

```

2837 TCACACGGCTGACGGTAACGGCAAAATGACNGTCAGAACATTCCGC 2886
      |||
      |||||
1089 erasn.....ValasnGlySerAlaasn..... 1096
2887 TTATATCGAACTCTTCGGCTACCGAACGACGACAAATTAAGCTGGCGA 2936
      |||||
1097 .....LeuIysValGlyTh 1101
2937 AAGTCCGAAGNACTTACACCTTGGCGGTCAACATACCGCAAGAAC 2986
      |||
      |||||
1101 rtrrProGluSerLysAspTyrGlyGlyLysPheAsnAlaGlyThrThn 1118
2987 CCCTAAGCCTCGATCAATTGACG.....GTACTGAAGGGA 3024
      |||
      |||||
1118 lshIsSerLysGluGlnThrThrAlaLysValGlyThrIleThrGlySer 1134
3025 GACAAACAAACCGCTGCCGAAAACCTTATTCACCTCCCAACAGCA 3074
      |||
      |||||
1135 GlnGlyIleGluLeuAsnAlaGlyHisAsnLeuThrLeuGlnGlyThrH 1151
3075 CGTCGATGCCGGCGCGCTGGCTTACCACTCATCCGCAAGAGCGGAGT 3124
      |||
      |||||
1151 sIeuSerSer.....GluGlnAspIleAla 1160
3125 TCCGCTCGATTAATCCGCTC..... 3144
      |||
      |||||
1160 euAsnAlaThrAsnLysValAspLeuGlnSerAlaSerSerGluHsth 1176
3145 ...AAGAACAGAGCTTCCGCAAACTCGCGCAAGCGCAAGAACCAAAA 3191
      |||
      |||||
1177 GlnLysGlyAsnAsnLeuSerGlyGlyValGlnAlaGlyPheGlyLys 1193
3192 ACAGCGGCAAAAAAGACAAACCGCAAGCCTTGACGGCGCTG..... 3231
      |||
      |||||
1193 smethr.....AspAspAlaSerSerValAsnGlyLeuGlySerAlaG 1208
3232 ..ATTGCGCGCGCGCGCGATGCCCGCAAAAGACAAACCGTTGCCGA 3279
      |||
      |||||
1208 lnpheAlaIleGlyLysGlnAspGlyLysSerValSerArgGlyGly 1224
3280 CCGGCGCGCGCGCGAGGC.....GGGGAATATGTCGGCAT 3314
      |||
      |||||
1225 ThrIleAsnAsnSerGlyAsnLeuThrIleAsnGlyAsnSerValHis 1241
3315 TATGCAGCGCGAGGAAGAAAAACGGTGCAGCGCGATAAGACAGCG 3364
      |||
      |||||
1241 uGlnGlyAlaGlnValAsnSerLysAspThrIleuThrSerGlnSerG 1258
3365 CMTTGGCGCAACAGCGCGCAAGCGGAACCGCGCGNTACACACCGCTTC 3414
      |||
      |||||
1258 lYAspIleGlyIleThrSerAlaGlnSerThrAspTyrLysAsnAsn 1274
3415 CCGCGCGCGCGCGCGCGCGGATTTGCCGCAACCGCGCAACCGCAAC 3464
      |||
      |||||
1275 GlyThrAspIleGlyPheAsnLysLysLysThrAsnAsnThrProLys 1291
3465 GCACCTCAACCCCAACCGCACCGCGACCTG..... 3495
      |||
      |||||
1291 uValThrGlnGlyLysProAlaThrSerIleHisAlaGlyGlyLys 1308
3496 .....ATNAGCCGTTATGCCAATAGCGGT 3519
      |||
      |||||
1308 euLeuValAsnValGlnAspGlnGlyLysThrSerHisGlnAsnAlaThr 1324
3520 TTGAGTGAATTTTCCGCGCAAGCGCTTTCGCGCGTAACAGAGAGA 3569
      |||
      |||||
1325 LeuGlyThrGlyThrLeuThrIleAsnSer.....AsnLysAs 1337
3570 ATTGACCGCGCTGTTGCCGAAGCGCGCGCAACCGNTTGGACAGACN 3619
      |||
      |||||
1337 pLeuThrLeuSerGlyAlaAsnValThrAlaAspSerVal...ThrGly 1353
3620 GCATCGCGNACACCAACACTACCTTCCGCAAGATTTCGCGCGCTACCGC 3669

```

```

1353 snValGlyGlySerLeuAsnIleAlaSerGln..... 1363
      |||
      |||||
3670 CAACAACCGGACCTGGCGCAAAATCGTATGCGAAGAAACCTCGCGACGG 3719
      |||
      |||||
1364 LysGluSerAsp...ArgHisVal..... 1370
3720 GCGCGTGGCATTCCTGTTTTCGCAACCGGACCGCAAAACNCTTCGACG 3769
      |||
      |||||
1371 .ThyValGlyValAsnValGlyTyrAsnHisThrAsnAspProLysSer 1387
3770 ACGGATCGGCAACTCGCGACGCGCTTGGCCAGCGCGCTTTTGGCGCA 3819
      |||
      |||||
1387 erGlnValAsnLysThrAlaLys...AlaGlyLysSerLeuGlnLys 1402
3820 TACGGCATCGCGAGTTGACATCGCATCGACACGCGCGCGGTTTAC 3869
      |||
      |||||
1403 ThrIleLysAspThrIleAspSerGlyLysSerSerThrAspAla 1419
3870 CAGC.....GGCANTCTNTCAGACGCGCATCGGAGGCAAAATCCGCC 3910
      |||
      |||||
1419 eSerAspLysTyrAsnSerLeuSerSerThrIleAlaAspLysThr... 1434
3911 GCCGCTGCTGCATTAACGCAAT.....CAGCACGATACCGC 3948
      |||
      |||||
1435 .....GlyIleSerAspGlyThrLysAlaLysIleAsp 1445
3949 GCCGCTTTCGCG.....GGATTCGCATCGCAACCGTATCGCGCGAC 3992
      |||
      |||||
1446 GlnGlyPheGlyLysValGlyAsnGlyIleLysAsnIleValThrGly 1462
3993 GCGCTATTGCTCCAAAAGCGGATTCACCGTACGAAACGTCAT 4038
      |||
      |||||
1462 aGlnGlyHisThrAlaAsnAlaAspIleLysValThrHisValAsp 1477
seq_name: SwissProt_40:AlDA_ECOLI
seq_documentation_block:
ID AlDA_ECOLI STANDARD; PRT; 1286 AA.
AC 003155;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Adhesin alda-I precursor.
GN AlDA-I.
OS Escherichia coli.
OG Plasmid p186.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.
RC STRAIN=2787 (O126:H27);
RX MEDLINE=92326638; PubMed=1625582;
RA Benz I., Schmidt M.A.;
RT "AlDA-I, the adhesin involved in diffuse adherence of the
RT diarrhoeagenic Escherichia coli strain 2787 (O126:H27), is
RT synthesized via a precursor molecule."
RL Mol. Microbiol. 6:1539-1546(1992).
CC -1- FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE
CC ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI
CC TO EPITHELIAL CELLS.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X65022; CAA46156.1; -

```

DR PIR: S28634; S28634.
 KW Cell adhesion; Signal; Outer membrane; Plasmid.
 FT SIGNAL 1 49 ADHESIN AIDA-I.
 FT CHAIN 50 ?
 FT PROPER 1286
 SO SEQUENCE 1286 AA; 132271 MW; B2A00F72AC05FB34 CRC64;

alignment_scores:
 Quality: 214.50 Length: 1456
 Ratio: 0.341 Gaps: 62
 Percent Similarity: 43.201 Percent Identity: 18.475

alignment_block:
 US-09-303-518D-651 x AIDA_ECOLI ..

Align seg 1/1 to: AIDA_ECOLI from: 1 to: 1286

```

646 TCCGGCGCATGGTTAATGGCGCAATACATATG.....CAGGGTTG 689
|||||: : : : : : : : : : : : : : : : : : : : : : : : : :
58 SerGlyGlyThrValSerSerGlyGlnThrGlnIleValTySerGlyYar 74
690 GGAATATATGGCGTANTAGTTTGACGGCGCATGTGCCCATGGCCACG 739
|||||: : : : : : : : : : : : : : : : : : : : : : : : : :
74 gGlyAsnSerAsnAlaThrValAsnSerGlyGlnThrGlnIleValAsn 91
740 ACTATGGCCCTATGCCGATTCGAGTGCAGCGCAGCGCACGGCTCGCCA 789
: : : : : : : : : : : : : : : : : : : : : : : : : :
91 sn...GlyGlySerThrThrAlaThrThrValAsnSerSerGlySerGln 106
790 ATGTTTATTTATGACAAACAAACAATAATGGCTGCTCAC..... 831
: : : : : : : : : : : : : : : : : : : : : : : : : :
107 AsnValGlyThrSerGlyAlaThrIleSerThrIleValAsnSerGlyG 123
832 .....GAGTTTACAAACCGGCTACCCCTTAT 859
|||||: : : : : : : : : : : : : : : : : : : : : : : : : :
123 yIleGlnArgValSerSerGlyGlyValAlaSerAlaThrAsnLeuSerG 140
860 CCGGCGAGGAGAAACGGTTTCACGCTGATACGAAAGATTGGTTTCAGAT 909
: : : : : : : : : : : : : : : : : : : : : : : : : :
140 yGlyValGlnAsnIleTyAsnLeu..... 148
910 GACATTACAGAGCGCATACATACCTCTCTTTTGAACCGCGCAGTAA 959
148 ..... 148
960 CGGACATTTTCTCTTACA.....TCCACACACACGGTAGC.... 996
|||||: : : : : : : : : : : : : : : : : : : : : : : : : :
149 .GlyHisAlaSerAsnThrValIlePheSerGlyGlnGlnThrIleP 165
997 .....GATACGCTACAGAACCCAGAAAGGTTTCCATCCAAAGCTT 1041
|||: : : : : : : : : : : : : : : : : : : : : : : : : :
165 heSerGlyGlyIleThrAspSerThrAsnIleSerSerGlyGlnGln 181
1042 AAGGTACAGACAGTCCGACTGTTTGACGAATCT..... 1074
: : : : : : : : : : : : : : : : : : : : : : : : : :
182 ArgValSerSerGlyGlyValAlaSerAsnThrThrIleAsnSerSerG 198
1075 .....TTGAATGAACCTGATPAAGAACGATTACGGCGGCG 1111
|||: : : : : : : : : : : : : : : : : : : : : : : : : :
198 yAlaGlnAsnIleLeuSerGlnGlnGlyAlaIleSerThrHisIleSer 215
1112 GGGGTGTATACAGTAC..... 1128
: : : : : : : : : : : : : : : : : : : : : : : : : :
215 erGlyGlyAsnGlnTyIleSerAlaGlyAlaAsnAlaThrGlnThrIle 231
1129 .....CGTCCAGGTTAAACACGCT..... 1149
|||: : : : : : : : : : : : : : : : : : : : : : : : : :
232 ValAsnSerGlyGlyPheGlnArgValAsnSerGlyAlaValAlaThrG 248
1150 .....GAAACCTTCTTTATCATGATTCAGTACGGCA 1177
: : : : : : : : : : : : : : : : : : : : : : : : : :
248 yThrValLeuSerGlyGlyThrGlnAsnValSer.....Serg 261

```

```

1178 ACGGCAAACTCATCTTATCAACACATCAACGAGCGGCGGCTTGG 1227
: : : : : : : : : : : : : : : : : : : : : : : : : :
261 yGlySerAlaIleSerThrSerValTyAsnSerGlyValGlnThrVal 277
1228 TATTTGAGAGT.....GATTTACGGTCTCCGCTTAAACAA 1265
|||: : : : : : : : : : : : : : : : : : : : : : : : : :
278 ...PheAlaGlyAlaThrValThrAspThrThrValAsnSerGlyGly 293
1266 CGAAACGTGGCAAGCGCGGCGTTCATATCAGTGAAGACAGTACGCTTA 1315
|: : : : : : : : : : : : : : : : : : : : : : : : : :
293 nGlnAsnIleSerSerGlyGlyIle...ValSerGlu.....Thr 306
1316 CTTCGAAAGTAAACGCGGTGCAAC.....GACCGCTG 1350
||: : : : : : : : : : : : : : : : : : : : : : : : : :
306 hrValAsnValSerGlyThrGlnAsnIleTySerGlyGlySerAlaLeu 322
1351 TCCAAAATCGCAAGGACGCTGCACGTTCAAGCCAAAGGGA.... 1395
|||: : : : : : : : : : : : : : : : : : : : : : : : : :
323 SerAlaAsnIleTySerGlnIleValAsnSerGlnGlyThrAlaI 339
1396 .....AACCAAGGCTCA 1408
339 eaSnThrLeuValSerAspGlyTyThrGlnHisIleArgAsnGlyTy 356
1409 TCACGCTGGCGACGCTACATCTTTGATCAGCAGCAGACAGATAA 1458
||: : : : : : : : : : : : : : : : : : : : : : : : : :
356 leAlaSerGly.....ThrIleValAsnGlnSerGlyTyValAsn 369
1459 GGCAAAACAGCCTTATGAAATAGCGCTGTGTCAGGCGGAGGCTAC 1508
370 IleSerSerGlyGlyTyAlaGlnSerThrIleIleAsnSerGlyGlyTh 386
1509 GTGCAACCTGAMTCCGATATATCATTCACCCGACAACTATTTGC 1558
||: : : : : : : : : : : : : : : : : : : : : : : : : :
386 rLeuArgValLeuSerAspGlyTyAlaArgGlyThrIleLeuAsn.... 401
1559 GCTTTCGGCGGACGCTTGGATTTAAACGGCATTCGCTTGGCTCCAC 1608
: : : : : : : : : : : : : : : : : : : : : : : : : :
402 ....AsnSerGlyArgGlnAsnValSerAsnGlyGlyValSerTyAsn 416
1609 CGTATTCAAATACCGAMTGAAGGGCGATGATTCNCATATCATATGCG 1658
||: : : : : : : : : : : : : : : : : : : : : : : : : :
417 AlaMetIleAsnThrGlyGlyAsnGlnTyIleTySerAspGlyGluAl 433
1659 AACACACATCCACCGTTACCATTCAGCG.....AATGAAGTA 1696
||: : : : : : : : : : : : : : : : : : : : : : : : : :
433 aThrAlaAlaIleValAsnThrSerGlyPheGlnArgIleAsnSerGly 450
1697 TTCACACACCG..... 1707
|||: : : : : : : : : : : : : : : : : : : : : : : : : :
450 yThrAlaPheValGlnAsnSerValValAlaThrArgThrValSerSer 466
1708 AGTGTAAGATATATCAATAGACTTAATTACAGC.....AAGAAATTCG 1751
: : : : : : : : : : : : : : : : : : : : : : : : : :
467 AlaAlaTySppPheAspAlaGlnValTySerGlyGlyGlnThrVa 483
1752 CTAC.....AACGTTGGTTGGCGAAGAAATAGC..... 1782
|||: : : : : : : : : : : : : : : : : : : : : : : : : :
483 lTyIleuThrArgGlyIleTyPyrSerAsnPheLeuThrAlaValTrps 500
1783 .....ACCAAAACGACGCGCG 1800
500 erMetPheProGlyThrAlaSerGlyAlaAsnValAsnLeuSerGlyArg 516
1801 CTCAAC..... 1806
|||: : : : : : : : : : : : : : : : : : : : : : : : : :
517 LeuAsnAlaPheAlaGlyAsnValValGlyThrIleLeuAsnGlnGln 533
1807 .....CTGTATACAGCCCGCGCA..... 1827
|||: : : : : : : : : : : : : : : : : : : : : : : : : :
533 yArgGlnTyValTySerGlyAlaThrAlaThrSerThrValGlyAsn 550

```



```
1828 ..GAGACGGCACCCNGCTGCTTTCCGGCGGAGCAAAATTAAAGCGCAAC 1875
      |||
      |||
550 snlgluylargylutryrvalleuserglylyllerhr...Aspglythr 565
1876 ATGACGAAACAAACGCAAACTGTTTACG.....GGCAGACCGAC 1919
      :
      :
566 ValLeuAsnSerIyglYleuGlnAlaValSerSerGlyIyLysAlaSe 582
1920 ACCGCAAGCCTTACATCATTTAGGAACGGGTGTGCAAAAATGGAAGTA 1969
      :
      :
582 rAlaThrValIleAsnGluGlyIyAla..... 591
1970 TCCCAACAGAGAAATCGTGTGGACACAGCTGATCNACCGCAGCTTT 2019
      :
      :
592 .....GlnPheValTyRAspGlyIyGlnValThrGlyThrAsn 604
2020 AAGCGGAAATTTCCATATTCAGGGCGGCGACGGCGGTATTTCCCGCA 2069
      :
      :
605 IleYAsnGlyIyThrIleArgValAspSerGlyAlaSerAlaLeuAs 621
2070 TGTTCGCAAAAGTGAAGCGGATTCATTTGAGCAAT..... 2106
      :
      :
621 nIleAlaLeuSerSerGlyIyAsnLeuPheThrSerThrGlyAlaThrL 638
2107 .....CACGCCAAGCAAGTTTGTGTGCGCACCGCATCAAAAC 2145
      :
      :
638 euProGluLeuThrThrMetAlaAlaLeuSerValSerGlnAsnHisAla 654
2146 CATCAATCTGT..... 2157
655 SerAsnIleValLeuGlnAsnGlyLeuLeuArgValThrSerGlyI 671
2158 .ACACGTTGCACTGACNGCTCTGCAAAATTTGTGCAANAANAACATTA 2206
      |||
      |||
671 yThrAlaThrAsp...ThrThrValAsnSerIaGlyArgLeuArgIleA 687
2207 CCGACGATAAAGTGTGCTTCACTGACTAAGACAGACNTNAGCGCANT 2256
      |||
      |||
687 spAspGlyIyThrIleAsnGlyThrThrIleAsnAlaAspGlyIle 703
2257 GTNAGNCTNNCNATNAGTNNNTNNAANCNCNGGCGNTGCNNACT 2306
      |||
      |||
704 Val.....AlaGlyThrAsnIleGlnAsnAspGlyIyAsnH 715
2307 NAANGCAATCTTACTGCAAAATGGGCAATACGTTATACAGTCAACCA 2356
      :
      :
715 eIleLeuAsnLeuAlaGluAsnTyRAspPheGluThrGluLeuSerGlys 732
2357 AGGCC.....ACCAAAACGGCAAC 2376
732 ergIyValLeuValLysAspAsnThrGlyIleMetThrTyRAlaGlyThr 748
2377 CTYAGCCTC.....GTGGCAATGCCCAAGCAACATTTAA 2411
      |||
      |||
749 LeuThrGlnAlaGlnGlyValAsnValLysAsnGlyIyIleIlePheAs 765
2412 TCAAGCCACATTAACGCAACNCATCGGNTCGGCAATGCTTCATTTA 2461
      :
      :
765 pSerAlaValAlaAsnAlaAspMetAlaValAsnGlnAsnAlaTyRleA 782
2462 ATCTAAGCAACAAGCGCGACAAAGGCAAGCTCAACGCTTCCGACAAAC 2511
      :
      :
782 snIleSerAspGlnAlaThrIleAsnGlySer.....ValAsnAsnAsn 796
2512 GCTAAGCAAAAGCTAAGCATTCGCACTCAACGCAATGTCCTCCTAGC 2561
      :
      :
797 GlySerIleValIleAsnAsnSerIleIleAsnGlyAsnIle..... 810
2562 CGATAAGCAGATTCATTTGAAAACAGCGGCTTTACCGGACAACTCA 2611
      |||
      |||
811 .....ThrAsnAspAlaA 815
2612 GCGGAGAGAGANACAGCATTAACATTAAGAGAGCAAGCAATGACGCTG 2661
```

```
815 sPLeuSerPheGlyThrAla.....LysLeu 823
2662 CCGTCAGCGACGGAATTAGCAATTTAAACCTTGACAAACGCCACCATTA 2711
      :
      :
824 LeuSerAlaThrValAsnGlySerLeu...ValAsnAsnLysAsnIleI 839
2712 ACTCAATTCGCCCTATCCGCACGATGCTGCGAGGCGCCCAACCGGCAAG 2761
      |||
      |||
839 eLeuAsn..... 841
2762 TGTACAGACAGCCGCGCGCGCTTCGGCGCTTCCTATTATCCGTTACA 2811
841 ..... 841
2812 CCGCCAACTTGGTAGAATCCGTTTCAACACGCTGACGTAACGGCA 2861
      :
      :
842 ...ProThrLys...GluSerAlaGlyAsnThrLeuThrValSer... 854
2862 ATTGAACGTCAGGAACATTCGCTTATGTCCGAACCTTCGCGTAC. 2910
855 ...AsnTyRThrGlyThrProGlySerValIleSerLeuGlyIyValL 870
2911 .....CGAAGCAAAATGGAAGCTGCGGCAAAAGTTCC 2943
870 euGluGlyAspAsnSerLeuThrAspArgLeuValIyLysGlyAsnThr 886
2944 GAAGGNACTTACACCTTGGCG...GTCAACAAATACGGCAACGAACCGT 2990
      :
      :
887 SerGlyGlnSerAspIleValTyRValAsnGlnAspGlySerGlyIyG 903
2991 AAGCCTGATCAATTAACGGTAGTGAAGGAAGCAACAAACCGCTGT 3040
903 nThrArgAspGlyIleAsnIleIleSerValGluGlyAsn..... 916
3041 CCGAAACCTTAATTTACCTTCGCAAAACGAACGCTGATGCGGCGCG 3090
      :
      :
917 ..SerAspAlaGluPheSerLeuLysAsn...ArgValAlaAlaGlyAla 931
3091 TGGCGTTACACATCCGCAAGACGCGAGTTCGCCCTGCATTAATCC 3140
      :
      :
932 TyRAspTyRThrLeu...GlnLysGlyAsnGluSerGlyThrAspAsn.. 946
3141 GGTCAAGAACACAGAGCTTCCGACAAACTCGGCAAGCAAGCAACAAA 3190
      |||
      |||
947 ...LysGlyTyRPTyRLeuThrSerHisLeuProThrSerAspThrArg 962
3191 AACAGCGGGAAGAACACAGCGCGCAAGCCTTGACGCGCTGATTCGCGC 3240
      :
      :
962 InTyRArgProGluAsnGlySerTyRAlaThrAsnMetAlaLeuAla... 977
3241 GGGCGCGATGCGCGCAAGAAAGCAAGAGCTTGCCGACCGCGCGCN 3290
977 ..... 977
3291 GGCAGCGGGGAAATGTCGCAATTAATGAGCGGAGAGAGAAAAAC 3340
      |||
      |||
978 .....AsnSerLeuPheLeuMetAspLeuAsnGluArgLysG 990
3341 GGGTCAGCGCGGATTAACACAGCGCNTTGGCGAAACACGCGCAAGCGAA 3390
      :
      :
990 InPheArgAlaMetSerAspAsn..... 997
3391 ACCGCGCGGNTTACACAGCGCTTCCCGCGCGCGCGCGCGCGGGA 3440
      :
      :
998 ThrGlnProGluSerAla..... 1003
3441 TTTCGCGCAACCGAGCCCAACCGCAACCTCAACCCCAACCGCAGCGG 3490
1003 ..... 1003
3491 ACCTGATNACCGTTATGCAATAGCGGTTTGAGTGAATTTTCGCGCAG 3540
```

```

1003 ..... 1003
3541 CTCACAGCGTTTGGCCGTACAGACGAATGACCGCGTTTGGCGA 3590
1003 ..... 1003
3591 AGACGGCGCACGCGNGTTGG.....ACAAGNCAGTCGCGNACA 3631
1004 ..... 1004
3632 CCAAACTACTACCGTTGGCAAGATTTCGGCGCTACCGCAACAAACGAC 3681
1016 lyls..... 1017
3682 CCGCGCAATGCGTATGAGAAAAACCTGGCGAGCGCGCGTGGCAT 3731
1018 ..... 1022
3732 CCGTTTTCACACACCGGACGAAACACTTCAGCAGCGCATCGCA 3781
1023 ..... 1036
3782 ACTCGGACGCGCTTGGCCGCGCGCGCTTTCGCGCATACGCGATCGC 3831
1036 lYAspIleTyrLysPheHisAlaGlu.....GlnLeuGly 1047
3832 AGGTTGACATCGGCATCGACAGCGCGCGGTTTACGACGCGGACNTCT 3881
1048 AspPheThrLeuGlyIleMetIleGlyTyrAlaAsnAlaIleGlyLys 1064
3882 NTCAGAGCGCATCGAGGCAAAATCCGCGCGCGTGTGATTCATACGCA 3931
1064 rIleAsnTyrThrSerAlaIleAlaArgAsnThrLeuAsp..... 1078
3932 TTCAGCAGCATACCGCGCGTTCGCGGATCGCATGCAACCGTAC 3981
1079 ..... 1092
3982 ATCGGCGCAAGCGGTATTTCGTCGCAAAA.....GCGGATTACCGTACGA 4028
1093 GlnAsnAlaThrGlyLeuPheAlaGluThrTrpMetGlnTyrAsnTrp 1109
4029 AAACGTCAATATCGCCACCGCGTGTGCGTTCAACGNTATCCG... 4074
1109 eAsnAlaSerValIleGlyAspGlyLeuGlnGluGlyTyrAsnLeuA 1126
4074 ..... 4074
1126 snGlyLeuThrAlaSerAlaGlyGlyTyrAsnLeuAsnValHisThr 1142
4075 ..... 4107
1143 TrpThrSerProGlnGlyIleThrGlyLeuPheTrpLeuGlnProHis 1159
4108 ..... 4153
1159 uGlnAlaValTrpMetGlyValThrProAspThrHis..... 1171
4154 ATGCGCGCTTGGGCAAGTCCGACACGCGGTCAATACCGCCTATTGGT 4203
1172 ..... 1179
4204 CAGGATTTCGCGCAAAACCGC.....AGGCGGA 4232
1180 GlnGlyAlaGlyLysAsnAsnIleGlnThrLysAlaGlyIleArgAla 1196
4233 ATGCGCGCTAAACCGCGCAATCAAGTTTCAGTGTCCNTCAACGCTG 4282
1196 rTrpLysValLysSer.....ThrLeuAspLysAspThrG 1208
4283 CCGCGCGCAAGGNCGCACTGGAAGCGAA.....CAC 4317
1208 lYArgArgPheArgProTyrIleGlnAlaAsnTrpIleHisAsnThrHis 1224

```

```

4318 AGCGCGCATCAATTA 4335
::: |||::|::|::|
1225 GlupheGlyValLysMet 1230
seq_name: SwissProt_40:AMYH_YEAST
seq_documentation_block:
ID AMYH_YEAST STANDARD: PRT; 1367 AA.
AC P08640: P08068:
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucosylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
DE glucosylase) (1,4-alpha-D-glucan glucosylase).
GN STAI OR STA2 OR MAL5 OR YIR0196.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (Dec-1994) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; PubMed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STAI."
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STA2 and STA genes
RT from Saccharomyces cerevisiae."
RL FEBS Lett. 239:179-184(1988).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: TO S.POMBE SPC215.13.
CC -1- SIMILARITY: SOME, TO S.POMBE SPC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z38061; CAA86176.1; -
DR EMBL: M16164; AAA55014.1; -
DR EMBL: M16165; AAA55015.1; -
DR EMBL: X13857; CAA32069.1; -
DR PIR: B26877; B26877.
DR PIR: A26877; A26877.
DR PIR: S48478; S48478.
DR SDD: S0001458; M0C1.
KW Hydrolyase; Glycosylase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1367
FT DOMAIN 210 1367
FT CARBOHYD 817 817
POTENTIAL.
SER/THR-RICH.
N-LINKED (GLCNAC. . .) (POTENTIAL).

```



```

3748 CGACGCGAAGAACCTGACGACGCGCATCGGACGACGCTTGC 3797
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
692 rla1proval1thrSerSerThrThlgluSerSerSerla...ProvalP 708
3798 CCAGGCGCGCTTTTCGGGCAATACGCGATCG..... 3829
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
708 roThr...ProSerSerSerThrThlgluSerSerSerlaProvalPro 723
3830 .....GACGCTTCGACATCGGACATACGACGCGGGGTTTTCGACG 3873
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
724 ThrProSerSerSerThrThlgluSerSerSerlaProvalProThrPr 740
3874 GGCATCTTTCAGACGCGATCGGACGCAAAATCCGCGCGCTGCTCA 3923
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
740 oSerSerSerThrThlgluSer.....SerSerlaProvalThrs 754
3924 TTACGCGATTCAGCAGCATACCGCGCGTTTCGGCGGATTGGCATCG 3973
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
754 erSerThrThlgluSerSerSerlaProvalProThrProSerSerSer 770
3974 AACCGTACATCGGCGCAACGCGCTATTTCGTCGCAAAAGGATTAACGC 4023
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
771 ThrThlgluSerSerSerlaProvalProThrProSerSerSerThr.. 786
4024 TAGCAAAACGTCATATCGCACCCCGGCTTTCGCTTCAACGCTACCG 4073
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
787 ThlgluSerSerSerlaProvalProThrProSerSerSerThrThng 803
4074 NCGCGCATTAAGCAGCAT.....ATTCAATCAACGCGCGC 4111
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
803 luserSerSerlaProvalProThrProSerSerSerSeranleThr 819
4112 AACACATTCATCAACNCCT.....ATTNAGC 4140
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
820 SerSerlaProSerSerSerThrProheSerSerSerThlgluSerSe 836
4141 CTGTCTATACGATCGCGCTTCGGGCAAGTCGACACGCGTCATATAC 4190
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
836 rAlProvalProThrProSerSerSerThrThlgluSerSerAlap 853
4191 CGCNGTATTGGCTCAGGATTTCGCAAAACCGGATGCGGATGGGCG 4240
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
853 ro.....ValSerSerSerThrThlgluSerSerAla 864
4241 .....TAACGCGCAATCAAG.....GTTTCAGCTGTCATTC 4275
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
865 ProvalProThrProSerSerSerSeranleThrSerSerAlaproSe 881
4276 CACGCTGCGC 4285
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
881 rSerlePro 884

```

seq_name: SwissProt_40:OMPb_RICTY

seq_documentation_block:
ID OMPb_RICTY STANDARD; PRT; 1645 AA.

```

AC P96989;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scad) (TompB)
DE antigen) (120 kDa outer membrane-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN OMPB OR SLP.
OS Rickettsia typhi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=785;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WILMINGTON;
RX MEDLINE=94040787; PubMed=8224886;

```

```

RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
RT "Cloning and sequence analysis of the gene encoding the crystalline
RT surface layer protein of Rickettsia typhi.";
RL Gene 133:129-133(1993).
RN [2]
RP PARTIAL SEQUENCE.
RC STRAIN-WILMINGTON;
RX MEDLINE=92114896; PubMed=1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNBr fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii.";
RL Mol. Immunol. 29:95-105(1992).
RN [3]
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE=92104668; PubMed=1729180;
RA Hackstadt T., Messer R., Cleplak W., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: identification of an avirulent
RT mutant deficient in processing.";
RL Infect. Immun. 60:159-165(1992).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMP/OMPb FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; I04661; AAB48987.1; -.
CC InterPro: IPR003858; TomPA; TomPB.
CC Pfam: PF02708; TomPA; TomPB; 1.
CC KW Antigen; S-layer; Transmembrane; Cell wall.
CC FT CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
CC FT TRANSMEM 1415 1645 32 KDA BETA PEPTIDE.
CC FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
CC FT CONFLICT 657 657 H -> N (IN REF. 2).
CC FT CONFLICT 842 842 V -> I (IN REF. 2).
CC FT CONFLICT 1071 1071 G -> A (IN REF. 2).
CC FT CONFLICT 1306 1306 G -> S (IN REF. 2).
CC SQ SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185BE CRC64;

```

alignment_scores:
Quality: 205.00 Length: 1203
Ratio: 0.390 Gaps: 54
Percent Similarity: 43.724 Percent Identity: 18.620

alignment_block:

US-09-303-518d-651 x OMPb_RICTY ..
Align seg 1/1 to: OMPb_RICTY from: 1 to: 1645

```

664 GCGCGCATAC.....CATATGAGGCTTGGGGAATATATG 701
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
259 GlyIysnThrIleasnPheasnGlyArgAspGlyThrIlylsLeuVa 275
702 CGTANTTAGTTGAGCGGCGATGCGCGCAATGCCAAGACTATGCGCCCTA 751
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
275 lLeuValSerIyasnGlysnAlaThrGluPheasn..... 287
752 TGCGATTGCGAGTGGCGAGCGCGATTCGCCAATGTTATTTAT 801
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
288 .....ValThGlySerLeuGlyIyasnLeuIyGlyValIleGluPhe 302

```

```
802 GACAAACA.....AACATTAATGCTGCTCAACGAGTTTACAAAC 845
      ||| ||| :||| :|||
303 AsphThrThrIleAlaIleGlyLeuIleAlaSnGly.....GlyAl 317
      :||| :||| :||| :|||
846 CGGCTACCTTATTCGGCAGGAAACGGTTCCAGCTGATACGCAAG 895
      :||| :||| :||| :|||
317 AlaAlaAlaValIleGlyThrAspAsnGlyAlaGly.....ArgAla 332
      :||| :||| :||| :|||
896 ATGCTTACGATGACATTTACAGAGCGGTACACATACCGTCTTTT 945
      ||| :||| :||| :|||
332 IagIlePheIleValSerValAspAsnGlyAlaAlaThrIleSerGly 348
      :||| :||| :||| :|||
946 GAACCGCGCAGTACGACATTTTCTTACATCAACAACAGCT.. 993
      :||| :||| :||| :|||
349 GluValThyAlaLysAspIleValIleGlnSerIleAlaSnIleGly 365
      :||| :||| :||| :|||
994 .....ACGGTACGGTACAGAAACAGC 1018
      :||| :||| :||| :|||
365 nValThrPheGluHisLeuValAspValGlyLeuGlyLysThrAsp 382
      :||| :||| :||| :|||
1019 AAAGGTTCATCCAAAGCTTAAGTACAGACAGTCCGACTTTGAC 1068
      ||| :||| :||| :|||
382 helThrAlaAspSerLysValIleIleThrGlu.....Asn 394
      :||| :||| :||| :|||
1069 GAATCTTTGAATGAAGTAT..... 1089
      ||| :||| :||| :|||
395 AlaserPheGlySerThrAspPheGlyAsnLeuAlaValGlnIleVal 411
      :||| :||| :||| :|||
1090 .....AAGAACCATTTACGGCGAGGGGGGTATATCAGTACCGTCAA 1135
      :||| :||| :||| :|||
411 LProAsnLysIleLeuThrGlyAsnPheIleGlyAspAlaLys.... 426
      :||| :||| :||| :|||
1136 GGTAAACACAGGTGAACCTTCTTTATCGATAC...GGCAGGCG 1182
      ||| :||| :||| :|||
427 .....AsnAsnGlyAsnThrAlaGlyAlaIleThrPheAsnAlaSnGly 441
      :||| :||| :||| :|||
1183 AAACATCTTATCAAC.....AACATCAA 1208
      ||| :||| :||| :|||
442 ThrLeuValSerGlyAsnThrAspProAsnIleValValThrAsnIle 458
      :||| :||| :||| :|||
1209 C.....CAAGCGCGGGCGGTGTGTTGTAAGGTATTTA 1246
      :||| :||| :||| :|||
458 AlaIleIleValGluGlyAlaGlyIleVal..... 468
      :||| :||| :||| :|||
1247 CGGTCTCGCTGAACAAACGAAACGTGCAAGCGCGGCTCATATC 1296
      ||| :||| :||| :|||
469 .....GlnLeuSerGlyIleHisGly 475
      :||| :||| :||| :|||
1297 AGTGA.....GACAG 1307
      :||| :||| :||| :|||
476 AlaGluLeuAlaGluGlyAsnAlaGlySerIlePheLysLeuAlaAsp 492
      :||| :||| :||| :|||
1308 TACGTTACTTGAAGAAAGTAAACGGCGTGGCAACGACCGCTGCCAAA 1357
      :||| :||| :||| :|||
492 yThrVal.....IleAsnGlyProValAsn..... 500
      :||| :||| :||| :|||
1358 TCGGCAAAAGCAGCGCTGACGCTTCAAGCAAAAGGGAACCAAGGCTCG 1407
      ||| :||| :||| :|||
501 .....GlnAsnProLeuValAsnAsnAla 509
      :||| :||| :||| :|||
1408 ATCAGCGTGGC.....GACGTAACATCTTTGATGCAGA 1445
      :||| :||| :||| :|||
510 LeuAlaIleGlySerIleGlnLeuAspGlySerAlaIleIleThrGlyAs 526
      :||| :||| :||| :|||
1446 GGCAGAGCATAAAGCAAAACAAAGCTTATAGTAATCGGCTGTC 1494
      :||| :||| :||| :|||
526 PileGlyAsnGlyAlaValAlaAlaLeuGlnAspIleThrLeuAla 543
      :||| :||| :||| :|||
1495 .....AGCGCAGGCGTACGCTCAACTG 1518
      :||| :||| :||| :|||
543 snAspAlaSerLysIleLeuThrLeuSerGlyAlaAsnIleIleGlyAla 559
      :||| :||| :||| :|||
1519 AATGCCATTAATCAAGTTCACCCGACAAACTATTTCCGCTTCCGG 1568
```

```
560 AsnAlaGlyGly.....AlaIleHisPheGlnAlaAlaSn 571
      :||| :||| :||| :|||
1569 CGGACCTTTGATTTA.....ACGGCATTCGCTTTGCTTC 1606
      :||| :||| :||| :|||
571 yGlyThrIleGlnLeuThrSerThrGlnAsnAsnIleLeuValAspPhe 588
      :||| :||| :||| :|||
1607 ACCGTAATCAAAATACGATGAAGGGCGATGATTCNCATCATTAATGCC 1656
      :||| :||| :||| :|||
588 splLeuAspValThrThrAspGlnThrGlyValValAspAlaSerSerLeu 604
      :||| :||| :||| :|||
1657 ACAACAATCCACCGCTTACCATTTACAGGAATGAAGTATTACAAACC 1706
      :||| :||| :||| :|||
605 ThrAsnAsnGlnThrLeuThrIleAsnGlySerIleGlyThrIleGlyAl 621
      :||| :||| :||| :|||
1707 GAGTGTAAAGATATCATATGACTTATATTCAGCAAA..... 1743
      :||| :||| :||| :|||
621 AsnThrLysThrLeuGlyArgPheAsnValGlySerSerLysThrIle 638
      :||| :||| :||| :|||
1744 .....GAATTCGCTCAACGCTTGGTTGGCAGAAAGAT... 1779
      :||| :||| :||| :|||
638 euAsnAlaGlyAspValAlaIleAsnGlnLeuValMetGluAsnAspGly 654
      :||| :||| :||| :|||
1780 .....ACGCCAAACG..... 1791
      :||| :||| :||| :|||
655 SerValHisLeuThrHisAsnThrTyLeuIleThrLysThrIleAsnAl 671
      :||| :||| :||| :|||
1792 .....AACGGCGGCTCAACCTTGTATACAGCCCGCGCAGAAAGAC 1834
      :||| :||| :||| :|||
671 AlaAsnGlnGlyLysIleIleValAlaAlaAspProIleAsnThrAsp 687
      :||| :||| :||| :|||
1835 GCACCCNGCTCTTCCGGCGAACAATTA..... 1866
      :||| :||| :||| :|||
688 .....ThrIleAlaAlaAspGlyThrAsnLeuGlySerAlaGluSerPro 702
      :||| :||| :||| :|||
1867 .....ACGGC..... 1872
      :||| :||| :||| :|||
703 LeuSerAsnIleHisPheAlaThrLysAlaAlaAsnGlyAspSerIle 719
      :||| :||| :||| :|||
1873 .....ACATCAGCAACAA 1888
      :||| :||| :||| :|||
719 uHisIleGlyLysGlyValAsnLeuTyAlaAsnAsnIleIleThrThA 736
      :||| :||| :||| :|||
1889 AC.....GGCAACTGTTTTCAGCGGAGACGACGACCGCAGCGC 1929
      :||| :||| :||| :|||
736 spAlaAsnValGlySerLeuHisPheArgSerGlyLysThrIleVal 752
      :||| :||| :||| :|||
1930 TACATCATTTAGGAACGGGTGCTCAAAATGGAAGTATCCACAGAG 1979
      :||| :||| :||| :|||
753 SerGlyThrValGlyGly.....GlnGlnGlyLeuLysLeuAs 765
      :||| :||| :||| :|||
1980 AGAAATCGTGTGGACACAGCAGTCAACCGGACGTTTAAAGCGGAA 2029
      :||| :||| :||| :|||
765 nasLeuIleLeuAspAsnGlyThrThrValAlaPheLeuGlyAspIle 782
      :||| :||| :||| :|||
2030 ATTTCATATTCAGGCGGAGCGGATTTCCGCAATGTGGCAAA 2079
      :||| :||| :||| :|||
782 hrPheAsn.....GlyGlyThrLysIleGluGlyLysSerIleLeuGln 796
      :||| :||| :||| :|||
2080 GTGGAAGCGATTCATTTGACCAATCAAGCCCAA..... 2115
      :||| :||| :||| :|||
797 IleSerSerAsn...TyrlleThrAspHisIleGluSerAlaAspAsnTh 812
      :||| :||| :||| :|||
2115 ..... 2115
      :||| :||| :||| :|||
812 rGlyThrLeuGluPheValAsnThrAspProIleThrValThrLeuAsn 829
      :||| :||| :||| :|||
2116 .....CGAGTTTGTGTGTGCGACCGCATCAAAAGCCATCAATCTGT 2157
      :||| :||| :||| :|||
829 yScIleGlyAlaArgPheGlyValLeuLysGlnValMetValSerGlyPro 845
      :||| :||| :||| :|||
2158 ACAGTTTCGAGCTGACNGCTGTGACAAATGTGTGGAANAACATTCAG 2207
      :||| :||| :||| :|||
```


GN OMPB OR RC1085.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_Taxid=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Mallish 7;
 RA MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samsen D., Roux V., Cossart P., Weissensbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
 RL Science 293:2093-2098(2001).
 RN [2]
 RP SEQUENCE OF 33-1649 FROM N.A.
 RC STRAIN-Indian tick typhus, and Mallish 7;
 RX MEDLINE=20393643; PubMed=10939649;
 RX Roux V., Raoult D.;
 RT "Phylogenetic analysis of members of the genus Rickettsia using the
 RT gene coding the outer-membrane protein rompB (ompB).";
 RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
 RN [3]
 RP SEQUENCE OF 353-1655 FROM N.A.
 RC STRAIN-Mallish 7;
 RA Stenos J., Walker D.;
 RT "The rickettsial outer membrane protein A and B genes of Rickettsia
 RT australis, the most divergent rickettsia of the spotted fever group.";
 RL Submitted (MAY-1999) to the EMBL/Genbank/DDAJ databases.
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
 CC SIMILARITY).
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 CC LAYER WITH HEXAGONAL SYMMETRY (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AE008659; AAL03623.1; -;
 DR EMBL: AF123721; AAF34124.1; -;
 DR EMBL: AF123726; AAF34129.1; -;
 DR EMBL: AF149110; AAD39533.1; -;
 DR InterPro: IPR003858; rompA, rompB.
 DR Pfam: PF02708; rompA, rompB; 1.
 KM Antigen; S-layer; Cell wall; Complete proteome.
 FT CHAIN 1 1334
 FT CHAIN 1 1334
 FT CHAIN 1335 1655
 FT CHAIN 1655
 FT VARIANT 61 61
 FT VARIANT 75 75
 FT VARIANT 78 78
 FT VARIANT 251 251
 FT VARIANT 413 413
 FT VARIANT 959 959
 FT VARIANT 988 988
 FT VARIANT 1139 1139
 FT VARIANT 353 353
 FT VARIANT 354 354
 FT VARIANT 776 776
 FT VARIANT 1159 1159
 FT VARIANT 1177 1177
 FT VARIANT 1492 1492
 FT VARIANT 1655 1655
 FT VARIANT 168342 MW; E49E19377D5FCB37 CRC64;
 FT SEQUENCE

alignment_scores:

Quality: 203.00 Length: 1500
 Ratio: 0.288 Gaps: 79
 Percent Similarity: 46.933 Percent Identity: 19.400
 alignment_block:
 US-09-303-518D-651 x OMPB_RICCN ..
 Align seg 1/1 to: OMPB_RICCN from: 1 to: 1655

```

178 AATAAGGCAAGTTTCACCTCGGGGGAAGATATTGAGTTTACACAA 227
   |||||
339 ASPASnglyLys.....ValaIatnrIleaspglyGlnValaItyrAlaLy 353
228 AAAAGGGGAGTTGTTCGCCAATCAATGACAAAAGCCGATATTGATT 277
   |||||
353 s.....AspmetValIleGlnSerIleasnAlaValaIglyGlnValaAsp 368
278 TTTCT.....GTGTGTGCGCTAACGGCGGTGGCGCATTTGGTGGC 318
   |||||
368 hearghisIleValaIspValaIglyThrAspGlyThrThrAlaIheysthr 384
319 GATCATATATTTGTGAGCGTGACATACGCGCGGTATACACAGTTGA 368
   |||||
385 AlaAlaSerLysValaIaIleThrGlnAsnSerAspIheGlyThrThrAs 401
369 TTTTGTGCGGGAAGAAAGNAATCCGATGACGCGTTTCTTACCAA 418
   |||||
401 pIhegly.....AsnleuAlaIaIleGlnI 409
419 TTTGTGAAGAAT.....AATTATAGCTTCACAT 450
   |||||
409 IeIleValProAsnThrmetThrLeuAsnGlyAspIheThrIleVala 425
451 TCACACCTTACACGCG.....GATTACATATGCG 482
   |||||
426 SerAspProGlyAsnThrAlaIleThrIleThrIleAspAlaAsn...GI 441
483 GCGTTGCGTAATTTGTCCAGATGACAGACCTGCGAAGAGAGAGT 532
   |||||
441 yThrIleuAlaSerAlaSerAlaIleAsn...ValaIaValIthrAsnA 457
533 ACATGAGGGGGAATACCTATTCGAT.....AAAGA 564
   |||||
457 snIleThrAlaIleGlnAlaSerGlyAlaIleValaIleGlnIleuSerGly 473
565 AATATCCCGAGCGTCTCCGATCGGCTGAGACACACACTATTGGCGTTA 614
   |||||
474 ThrHisAlaIleGlnIleuArgIleuGlyAsnAlaIleGlySerValPheLysIle 490
615 TGATGATGACAAA.....CACGGGATTTATCTACTCGGCGCATGAT 658
   |||||
490 uIaIaAspGlyThrIleValIleAsnGlyLysValaIleGlnThrAla.....L 505
659 TAAATGGCGGCAATACACATATGACAGGTTGGGGAATATATGCGCATNT 708
   |||||
505 euValaIleGlyAlaIleuAlaIaIleGlyThrIleThrLeuAspIleGlyAla 521
709 AGTTGAGGGGAGATGTCGCCATGCAACGACTATGGCCCATATGCCAT 758
   |||||
522 ThrIleThrGlyAspIleGlyAsnAla..... 530
759 TGCAGGTGCGGAGGAGACAGCGGTTCGCAATGTTATTATGACAAA 808
   |||||
531 .GlyGlyAlaIleAlaIleGlnGlyIleThrIleu...AlaAsnAspAlaIat 546
809 CAACATATATATGCTGCTCAACGAGATTTCACAAACGGCTACCCCTTAT 858
   |||||
546 hrlYsthrIleuThrIleu..... 551
859 TCCGCGAGGGAACGAGTTCCAGCTGATACGCAAGAATGTTGTCTACGA 908
   |||||
552 .....GlyGlyAlaAsnIleIleGlyAlaAsn.GlyGlyThrI 564

```


1077GlyAsnIleGlyAspSerAspThrP 1085
2513 CTAAAGCAACGTAAGCCATTCGACCTCAACGCAATGTCTCCAGCC 2562
1085 roValAlaSerValArgPheThrGlySerAspGlyValAlaGlyLeu... 1100
2563 GATAGGCAAGTATTCATTGTAAGAAACAGCCGCTTATCCGCAACCTCAG 2612
1101GlnGlyAsnIleTyrSerGlnValI 1109
2613 CGGCACACGAGNACGATTAACCTTAATAAGACGCAATGAGCGTGC 2662
1109 eAspPheGlyThrTyrAsnLeuGlyIleSerAsnSerAsnValIleLeuG 1126
2663 CGTACAGCAGCAAGATTA.....GGCAATTTAAACCTTACACACGCGCAC 2706
1126 LyGlyGlyThrThrAlaIleAsnGlyLysIleAsnLeuArgThrAsnThr 1142
2707 ATTACACTCAATTCGCTATGCGACAGATGCTGACGCGCGCAACCGG 2756
1143 LeuThrPhe.....AlaSerGlyThrSerThrTr 1152
2757 CAGNGTGTGACAGCGCGCGCGCTTCGCGCGCTTCCCTATATTCG 2806
1152 pGlyAsnAsn..... 1155
2807 TTACACCGCCCACTTCGTGAGATCCGCTTTCACACGCTGACGGTA... 2853
1156ThrSerIleGlyThr.....ThrLeuThrLeuAla 1165
2854 AACGGCAATTTGACNGTCAGGAACATTCGCTTATGTCGGAACCTGT 2903
1166 AsnGlyAsnIle.....GlyAsnIleValIleLeu..... 1175
2904 CGGCTACGCAACGCAATTTGACCTGCGGAAGTCCGAGNACTT 2953
1176GlnGlyAlaGlnValAsnAlaThrThrThrGlyThrT 1188
2954 ACACCTTGCGGCTCAACATACCGCAACGAAACCGTACGCTGATCAA 3003
1188 hThrIleLysValGlnAspAsnAlaAsnAlaAsnPheSerGlyThrGln 1204
3004 ...TTGACGGTGTGTGGAAGG.....AAAGCAACAACCGCTG...TC 3041
1205 ThrTyrThrLeuIleGlnGlyAlaArgPheAsnGlyThrLeuGlyG 1221
3042 CGAAACCTTATTTACCCCTGCAAAAGACAGCTGATCCGCGCGGT 3091
1221 yProAsnPheValIleThrGlySerAsnArgPheValAsnTyrGlyLeuI 1238
3092 GCGGT.....TACCAACTCATCCGCAAAAGACGCGAGTTTC 3126
1238 leArgAlaAlaAsnGlnAspTyrValIleThrArgThrAsn..... 1251
3127 CGCCTGATTAATCCGCTCAAAAGACAGCTTCCGCAAACTCGGCAA 3176
1252 AsnAlaGlnAsnValIleThrAsnAspIleAlaAsnSerSerPheGlyG 1268
3177 GGCAGAACCCCAAAAACAGCGGCAAAAAGACAGCGCAAGCCGTGACG 3226
1268 yAlaProGlyVal.....GlyGlnAsnValIleThrT 1278
3227 CGCTGATTTGCGCGCGCGCGATGCGCGCAAAAAGACGAAAGGTTGCC 3276
1278 hPheValAlaAlaThrAsnThrAlaAlaIleTyrAsnAsnLeuLeuAla 1294
3277 GAACGCGCGCGCGCGCGAGGGGGAAT...GTCCGATTTATGACAGC 3323
1295 Lys.....AsnSerAlaAsnSerAlaAsnPheValGlyAlaIleValThr 1309
3324 GGAGAGAGAGAAAAACGGGTGACGCGGATTAAGACAGCGCCTTTGGCGA 3373
1309 rAspThrSerAlaAlaIleThrAsnAlaGlnLeuAspValAlaLysAspI 1326
3374 AACAGCGCGAGCGGAAAAACCGCGGNTAACACGCGCTTCCCGCGGCC 3423
1326 leGlnAlaGlnLeuGlyAsnArg..... 1333
3424 CGCNCGCCCGCGGATTTGCCCAACCGCAGCGCCCAACGCAACCTCA 3473
1334 LeuGlyAlaLeuArgTyrLeuGlyThrProGluThrAlaGlnMetAlaG 1350
3474 ACCCAACCGCAGCGGACCTGATNAGCCGTTATGCCAATAGCGTTTGA 3523
1350 yPro..... 1351
3524 GTGAATTTCCGCGACGCTCAACAGCGTTTCCGCTACAGAGCAATTCG 3573
1352 ..GlnAlaGlyAlaIleProAlaAlaValAlaAlaGlyAspGlnAlaVal 1367
3574 GACCGCGTGTTCGCCGAAGACCGCGCAAGCGGNTTGGACAAGCNGCAT 3623
1368 AspAsnVal.....AlaTyrGlyIleThrAlaLysProPh 1379
3624 CCGGNACACCAACACTAC...CGTTGCAAGATTTCCGCGCTACCGCC 3670
1379 eTyrThrAspAlaHisGlnSerLysGlyGlyLeuAlaGlyTyrLysA 1396
3671 AACAAACCGACCTTCGCCAAATCGTATGCAGAAAAACCTCGCAGCGGG 3720
1396 lAlysThrThrGlyValIleGlyLeu..... 1405
3721 CGCGTCGCAATCCGTGTTTGCACACCGGACGAAACANCTTCACGA 3770
1406As 1406
3771 CGGATGCGCACTCGCGACGCTTGCCACGCGCGCTTTTGGGCAT 3820
1406 pThrLeuAlaAsnAspAsnLeuMetIleGlyAlaAlaIle..... 1419
3821 ACGCATCGGAGGTTCGACATCGGCATACGACGCGCGCGGCTTTTACG 3870
1420 ..GlyIleThrLysThrAspIle.....LysHisGlnAspTyrLys 1432
3871 AGCGGCAANTCTCAGACGCGATCGAGGCAAAATCCGCGCGCGTGTCT 3920
1433 LysGlyAspLysThrAspValAsnGlyPheSerPheSerLeu..... 1446
3921 GCATTTAGCGCATTCAGCAGCA.....TACCGCGCGGTTTCG 3958
1447TyrGlyAlaGlnGlnLeuValLysAsnPhePheAlaGlnGlySerA 1462
3959 GCGGATTCGCGCATCGACCGTACATCGGC...GCAACGCGCTATTTCGTC 4005
1462 lAllePheSerLeuAsnGlnValLysAsnLysSerGlnArgTyrPhePhe 1478
4006CAAAAAGCGATTTACCGCTACGAAAA 4031
1479 AspAlaAsnGlyAsnMetSerLysGlnIleAlaAlaGlyHisTyrAsps 1495
4032 CGTCAATATGCCACACCCCGCTTTCGCTTCAACCGCTACCGGCGGCA 4081
1495 nMetThr.....PheGlyLysnL 1502
4082 TTAAGCGAGATTTATTCATTAACCGCGCAACACATNTTCATCAGCCT 4131
1502 euthrValGlyTyrAspTyrAsnAlaMetGlnGlyAlaLeuValThrPro 1518
4132 TATTTNAGCTGTCTATACCATGCGCGCTTCGGGCAAAAGTCGAACAG 4181
1519 MetAlaGlyLeuSerTyrLeuLysSerSerAspGlnAsnTyrLysLunh 1535
4182 CGTCAATACCGGNGTATTCGCTCAGGATTTGGCAAA...ACCCGAGTG 4228
1535 rGlyThrThrValAlaAsnLysGlnValAsnSerLysPheSerAspArgT 1552


```
220 euProAspSerGlyIleSerValAlaLeuGluAspAsnThr..... 233
1321 AAGTAAACGGCGTGGCAACGACCGCTGTCCAAAATCGGCAAGCAC 1370
233 ..... 233
1371 GCTGCACGTTCAAGCCAAAGGGAAACCAAGGCTCGATCAGCGTGCGG 1420
234 .....ValGInGlyIleValLysAsnGInGlySerIleLysAlaGlyG 248
1421 ACGGTACAGTCAATTTGGATCAGCAGCGACAGATAAGCAAAAAACAA 1470
248 IuIleThrLeu.....SerAlaLysGlyArgLysGln 258
1471 GCCTTAGTGAATGCGCTGTGTCAGCGC..... 1500
259 AlaLeuAspSerLeuValMetAsnAsnGlyValLeuGluAlaThrLysVa 275
1501 .....AGGGTACGCTGCAACTGTAATGCGATTAATCAGTTCAACC 1540
275 LserAsnLysAsnGlyLysValValLeuSerAlaAspAsn..... 288
1541 CCGAAACACTCAATTTGCGCTTTCGCGCGACGCTTTGGATTAAACGG 1590
289 .....ValGluLeuAsnAsn 293
1591 CATTCGCTTTGCTTCACCGTATTCAAAATACGATGAAGGGGATGAT 1640
294 GluSer.....AsnIleLysGlyIuI 301
1641 TGCNATCATATATGCCACACACATCCACGTTACCATACAGGAGATG 1690
301 eValThrPheGlyAlaAspValThrSerAsnLysGluLeuLysAspAsnI 318
1691 AAGATATACAAACCGAGTGTAGAATATCAAT.....AGACTTAAT 1734
318 IeLysIleThrSerLysThrGlySerValThrSerProLysIleAsn 334
1735 TACAGC...AAGAATTTGCTTACACAGGTTGTTGGCGAGAAGAATAC 1781
335 PheThrGlyLysSerValAsnIleAsnGlyAsnPheGlyArgGluAspSe 351
1782 GACC.....AAACGACGCGCGGCTCAAC 1807
351 tThrThrHisTyrLysAspGluPheLysLeuAsnThrGluValAsnI 368
1808 TT.....GTTTACACGCGCGCGACGACGACGCGC 1836
368 IeAspValProAspAsnGluAsnIleArgIleAlaAspIleGluAspAsn 384
1837 ACCCGTGTCTTCCGGCGGAACA..... 1860
385 ThrGlyThrGlyThrThrGlyThrGlyThrSerSerPheIleGlnThrG 401
1861 .....AATTAAACGCA 1873
401 yAlaLeuSerSerLeuLeuAlaAsnAsnGlyLysValAsnLeuLysGly 418
1874 AC..... 1875
418 snAsnValAsnIleSerGlyArgIleHisIleAspSerPheArgGlySer 434
1876 .....ATCAGCAACAAAC..... 1890
435 AspSerLeuLeuLysLeuThrAsnLysGlyHisIleAspIleAsnAl 451
1891 .....GGCAACGTGTTTC..... 1905
451 aaSPHeHisSerLysGlyLeuPhePheIleThrSerLeuGlnAsnG 468
1906 .....AGCGGACAGACGACGACGACGCGCTCAATCATTTA 1941
468 IuGluAspPheLysSerAsnIleThrIleThrAspSerLysIleAsnLeu 484
1942 GGAACCGC..... 1950
485 GlyAsnGlyAlaMetGlyLeuGlyArgSerValAspGluLysAspTyrAs 501
1951 .....TGCTCAAAAATGGAAGT.....ATCC 1972
501 pasnArgTTPGInLysThrGluGlySerGlnArgLysLysPheAspValL 518
1973 CACAAGGAAATCGTGGGACAC..... 1998
518 ysmetSerAsnValGluPheAsnGlnValAspAspValIleLeuAlaGly 534
1999 .....GACTGATCMACCGACGTTTAAACGCA 2027
535 GlyPheGluLysValAsnLeuAspLysIleValAlaThrGlyGlnThr.. 550
2028 AATTTCCATATTCAGGCGGCGGCGGATTTCCGCATGTTGCCA 2077
551 .AsnPheTyrIleAspGly.....ValSerArgAsnGlyArgL 564
2078 AAGTGA...GGCGATTGNCATTTGAGCAATCAGCCCAAGCA..... 2118
564 ySTyrGluTyrGlyValLeuAspLeuAspLysArgThrGlnLeuSerGlu 580
2119 .....GTTTTGCTGTCGACCGCA 2138
581 LeuAsnGlnGlyArgArgArgTyrGlyTyrTyrAspLeuGluLeuAs 597
2139 TCAAGCATACATCTGTATACAGTTCGGACGTCGACNGGTGACAAT 2188
597 pMetAsnArgAlaTyrLeuTyrArgPheAspLeuPheAlaThrLysAsnT 614
2189 GTGTCAANAANCAATTACCGACGATAAGTG..... 2220
614 hrcLysArgSerThrIleLysAspThrGluIleAsnIleSerAsnSeran 630
2221 .....ATTGCTTCATGACTTAAGC 2240
631 IleAsnLeuLysAsnGlyPheValHisLeuLeuAlaGluLysIleLysLe 647
2241 NGACNTNAG.....GGCANTGTA 2260
647 uAspAsnSerLysIleAspIleThrPheAspLysAspAsnSerGlnAspT 664
2261 GNCNNCCNATNACGNTNNTNAAANCCTCNGCG...CNTCCNMCCTN 2307
664 hrcLeuAlaGlnThrAsnArgLeuGlyMetAsnGlyLysValSerMetIle 680
2308 AANGCAATCTTAGTCAATGAGCGATACAGCTTATACAGTCAGCACAA 2357
681 AsnSerHisIleLysIleValGlyAspGluLysGluGlyIleSerProh 697
2358 CGCCACCCAAACGGCACCTT.....AGCTCGTGGGCAATGCC 2398
697 rcGlyThrTyrAlaThrMetPheLeuIleGlyGluLeuIleGlyGluLys 714
2399 AAGCAACATTTAATCAAGC.....ACATTAACGCG 2430
714 eTserIlePheValLysSerHisGlnGlyTyrThrPheLysThrAspGly 730
2431 AACNCATCGNNTTCGGCAATGCTTCA..... 2457
731 AsnThrLysIleAlaGlyLysTyrSerLysGluAspLeuLysIleThrAl 747
2458 .TTTAATCTAAGCAACACGCGCACAA.....AACGCGAGTC 2494
747 aIleAsnThrGlyGlyArgAlaAlaGlnGluValLeuIleAsnGlyAlaL 764
2495 TGACGCTTCGCAACAGCTAAGGCAACGTAAGCATTCGCACTCAAC 2544
764 euGlySerAlaAspAsn...AspAlaAsnIleAlaAsnMetAlaPhe... 778
```

```

2545 GGCAATGTCCTCCAGCCGATAGCAGATATTCATTTCGTAACACCGC 2594
779 .....ThrliegIyaspSerAla.....AsnThrly 787
2595 CTTTACCGGACACTCAGCGGCGACAGANACGATTCACCTTAAAG 2644
787 sThrThr.....IlegIuAsnAlaAspIleThrAlaLeu..... 798
2645 ACAGGATGAGCGCTGCCGACGACGAGATTAATGCAATTTAACTT 2694
799 .....AlaProAsnGly.....GlyThrAlaTyrLeu 807
2695 GACAACGCCACCATTAACACTC.....AATTCGCCCTATCCGCA 2732
808 SerSerIyaspValGluIleGluValLysProAsnSerAsnPe..... 822
2733 CGATGCTCAGAGCGCGCAACCGGAGNGTGCAGACACCGCGCGCC 2782
823 .....ThrPheGluLeuProArg..... 829
2783 GTTCGCGCGCTCCCTATTATCCGTACACCGCACTTCGTGATTC 831
830 .....GluLys 831
2833 GCTTCAACAGCTGACGCGTAACGCG.....AATGACNGTCA 2873
832 AsnLeuAsnGlnThrLysIleAsnGlyAlaSerThrLysLeuSerGluAr 848
2874 AGGAACATTCGCTTTATGTCGGAACCTTCGCTACCGACGACCAAT 2923
848 gGlyPheAlaIleGlyLeuTyrAspLysIleAsnGlyValAlaIleSerAsn 865
2924 TGAACCTGGCGGAAGTCCGAGAGNACTACACTTCGCGGTCAACAT 2973
865 eu..... 865
2974 ACCGGCAACGAACCCGTAAGCTCGATCATTAAGTAGTGAAGGAA 3023
866 .....SerAlaGluGlnIleuAsnValThrAspAlaSe 876
3024 AGACACAACCGCTGTCGGAACCTTAATTCACCTCGCAACACGAC 3073
876 rGlu.....LysIleIleAsnThrLysLeuValSerSerLeuAspValGlu 892
3074 ACGTC.....GATCGCGCGCGTGCCTTACCAACTC 3105
892 yslLeuValSerValAlaValLysAspAlaGly..... 902
3106 ATCCGCAAGAGCGGCGAGTTCGCTGATATATCCGTCACAAAGACA 3155
903 .....AsnGlyCysGluGlnGlnI 909
3156 GCTTCCGACAAA.....CTCGCAACGACGAGAG 3184
909 nPheGlyAspLysGlyAsnAsnThrLysValSerValGlyLeuGlnA 926
3185 CCAAAAA 3192
926 laGluGln 928
seq_name: SwissProt_40:FRPC_NEIMC
seq_documentation_block:
ID FRPC_NEIMC STANDARD; PRT; 1829 AA.
AC P5127;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Iron-regulated protein frpc.
GN FRPC.
OS Neisseria meningitidis (serogroup C).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=135720;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FAM20 / SEROGROUP C;
RX MEDLINE=94018616; PubMed=8412674;
RA Thompson S.A., Wang L.L., Sparling P.F.;
RT "Cloning and nucleotide sequence of frpc, a second gene from
RT Neisseria meningitidis encoding a protein similar to Rtx
RT cytotoxins."
RL Mol. Microbiol. 9:85-96(1993).
CC -!- FUNCTION: MAY PARTICIPATE IN THE PATHOGENESIS OF MENINGOCOCCAL
CC DISEASE.
CC -!- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED AND SECRETED.
CC -!- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
CC CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
CC ACTIVITY.
CC -!- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L06299; AAA99902.1; -
DR InterPro: IPR001343; HemIysn_Ca_bind.
DR Pfam: PF00353; hemolysinCbind; 9.
DR PRINTS: PR00313; CARNDNGRPT.
DR PROSITE: PS00330; HEMOLYSIN_CALCIUM; 17.
KW Toxin; Calcium; Outer membrane; Repeat.
FT DOMAIN 879 1702 43 X REPEATS, GLY-RICH.
FT REPEAT 879 884 1.
FT REPEAT 888 893 2.
FT REPEAT 897 902 3.
FT REPEAT 1016 1021 4.
FT REPEAT 1025 1030 5.
FT REPEAT 1034 1039 6.
FT REPEAT 1043 1048 7.
FT REPEAT 1052 1057 8.
FT REPEAT 1061 1066 9.
FT REPEAT 1070 1075 10.
FT REPEAT 1079 1084 11.
FT REPEAT 1088 1093 12.
FT REPEAT 1097 1102 13.
FT REPEAT 1216 1221 14.
FT REPEAT 1225 1230 15.
FT REPEAT 1234 1239 16.
FT REPEAT 1243 1248 17.
FT REPEAT 1252 1257 18.
FT REPEAT 1261 1266 19.
FT REPEAT 1270 1275 20.
FT REPEAT 1279 1284 21.
FT REPEAT 1288 1293 22.
FT REPEAT 1297 1302 23.
FT REPEAT 1416 1421 24.
FT REPEAT 1425 1430 25.
FT REPEAT 1434 1439 26.
FT REPEAT 1443 1448 27.
FT REPEAT 1452 1457 28.
FT REPEAT 1461 1466 29.
FT REPEAT 1470 1475 30.
FT REPEAT 1479 1484 31.
FT REPEAT 1488 1493 32.
FT REPEAT 1497 1502 33.
FT REPEAT 1616 1621 34.
FT REPEAT 1625 1630 35.
FT REPEAT 1634 1639 36.
FT REPEAT 1643 1648 37.
FT REPEAT 1652 1657 38.
FT REPEAT 1661 1666 39.
FT REPEAT 1670 1675 40.
FT REPEAT 1679 1684 41.

```



```
1324 .....GTAAACGGCGTGGCAACGACCGCTGTCCAA 1355
1325 LysAspIleLeaRphetrAspGlyIleThrAlaAspMetLeuThr 1341
1356 AATCGCAAAAGGCAGC...CTGCACGTTCAACCAAGGGAACCAAG 1402
1341 eThraGluGluYAsnHisLeuLeuLeuLysAlaLys...AspAspSerG 1357
1403 GCCTCATCAGCGTG.....GGCGACGGTACAGTC 1431
1357 LysIleValThrValGlnSerTyrPheGlnAsnAspGlySerGlyAlaTyr 1373
1432 ATTTTGATCAGCAGCAGCATTAAGCAAAACCAAGCCTTAGGA 1481
1374 ArgIleAspGluIleHisPheAspAsnGlyLys.....Va 1385
1482 AATCGCTTGNTCAGCGGCGGATACGGTGCACACTGAATGCCGTAATC 1531
1385 IleuAspValAlaThrValGluLeuValGlnGlnSerThrAspGlyLys 1402
1532 AGTTCAACCCGCAAACTCTATTTCGGCTTCGGCGGCGAGCTTGAT 1581
1402 et.....AspArgLeuTyr...AlaTyrGlnSerGlySer...Thr 1413
1582 TTAAACGGCGCATTCGCTTCGCTTCACCGTATTCAAAATACCGATGAAG 1631
1414 LeuAsnGly...GlyLeuGlyAspAspTyrLeuTyrGlyAlaAsp...G 1428
1632 GCGCATGATGNCATCATATGCCACAACAATCCACCGTTACCATTA 1681
1428 YAspAspLeuLeuAsnGlyAspAlaGlyAsnAspSerIleTyrSerGlyA 1445
1682 CAGGAATGAAGATATTACACACCGGATGTAAGATATACATATGACTT 1731
1445 snGlyAsnAspThrLeuAspGlyGlyGlyGly...AsnAspAlaLeuTyr 1460
1732 AATTACACCAAGAAATTGCTTACACCGGTTGGTGGCGAAGAAAGATAC 1781
1461 GlyTyrAsnGlyAsnAspAlaLeuAsnGlyGlyGly..... 1473
1782 GACCAAAAGCAAGCGGCGTCAACCTGTTTACCAGCCCGCCAGAG 1831
1474 .....AsnAspHisLeuAsn.....GlyIleuA 1481
1832 ACCGACCCNGCTGCTTTCGCGGAAACAATTTA..... 1866
1481 spGlyAsnAspThrLeuIleGlyGlyAlaGlyAsnAspTyrLeuGly 1497
1867 ...AAGCGACATCACCAGCAACAAGGCAAACTGTTTTCACGGCAG 1913
1498 GlySerGlySerAspThrTyrValPheGlyLysGlyPhe.....GlyG 1512
1914 ACCGACACCGCAGCGCTTACATCATTTAGGAAGCGGCGTCAAAATG 1963
1512 nasPThrValTyrAsnTyrAspTyr..... 1520
1964 AAGGATATCCCAAGAGAAATGCTGTGGCAACGACTGATCACCAGC 2013
1521 ..AlaThrGlyArgLysAspIleLeaRphetrAspGlyIleThrAla 1536
2014 .....ACGTTTAAAGCGAAATTTCCATATTTCAGGGCGGCGAGC 2054
1537 AspMetLeuThrPheThrArgGlyAlaHisLeuLeu..... 1549
2055 GGTGATTTCCCGCAATGTGGCAAGTGAAGCGGATTCGATTTGAGCA 2104
1550 .....IleLysAlaLysAspGlySerGlyGlnValThr. 1560
2105 ATCAGCGCCAAGCAGTTTGGTGTCCACCGCATCAAGCATACATC 2154
1561 .....ValGlnSerTyrPhe..... 1565

2155 TGTAACGTTTGGACTGACNGCTGTGACAAATTGTGTGCAANAANCAAT 2204
1566 .....GlnAsnAspGlySerGlyAlaTyrArgLysAspGluIleHisP 1580
2205 TACGAGATTAAGTG.....ATTGCTCATCTACTAAG..... 2238
1580 eAspAsnGlyLysValLeuAspValAlaThrValLysLysLeuValGlnG 1597
2239 .....ACNGACNTNAGCGCANCTGTNAGCTNNCANTNACGNTNNTTNA 2283
1597 InsertThrAspGlySerAspArgLeuTyrAlaTyrGlnSerGlyAsnThr 1613
2284 AANTCTCNGCGCTGCN.....NCACNTAANAGC..... 2313
1614 LeuAsnGlyGlyLeuGlyAspAspTyrLeuTyrGlyAlaAspGlyAspAs 1630
2314 .....AATCTTATGCAATGCGCATTAACGTTATACGTTATACGCCACA 2356
1630 PLeuLeuAsnGlyAspAlaGlyAsnAspSerIleTyrSerGlyAsnGlyA 1647
2357 ACGCCACCCAAAC.....GGCACTTAGCTCTGTGGCAATGCC 2397
1647 snAspThrLeuAsnGlyGlyGlyGlyAsnAspAlaLeuTyrGly..... 1661
2398 CAAGCAACATTATATCAGCCACATTAAACGGCAACNCATCGGNTTCGGG 2447
1662 .....TyrAsnGlyAsnAspValLeuAsnG 1670
2448 CAATGCTTCATTATTAATCTAAGCAACAACGCCGCAAAAGCGAGCTCA 2497
1670 YGlyGlyGlyAlaAsnAspHisLeuAsn.....GlyGlyAspGlyAsnAsp 1685
2498 CGCTTTCGCAACGCTAAGGCAACGTAAGCCATTCCGCACTCAAGGCG 2547
1685 ThrLeuIleGlyGlyAla.....GlyAsnAspTyrLeuGlyGly 1697
2548 AATGTCCTCTACCGATTAAGCGATATTC..... 2577
1698 Gly...SerGlySerAspThrTyrValPheGlyLysGlyPheGlyGlnAs 1713
2578 .....CATTTGAANAACAGCGCTTACCAGCAACATCA 2611
1713 rPhrValTyrAsnTyrHisValAspLysAsnSerAspThrMetHisPhe 1730
2612 GCGCGACGAAGANACGATTTACAC...TTAAACACAGCGCAATGGAGC 2658
1730 YsGlyPheLysAlaAlaAspValHisPheIleTyrSerGlySerAspLeu 1746
2659 CTGCCGTCAGGCAAGGAAATTAGCAATTTAACTT..... 2694
1747 ValLeuSerAlaSerGlyGlnAspAsnValArgIleSerGlyPhePheTyr 1763
2695 .....GACAACGCCACCATTA 2710
1763 rGlyGlnAsnHisArgValAspThrPheValPheAspAspAlaIleHis 1780
2711 CACTCAATTCGCGCTATCGCACGATGCTGACAGC..... 2745
1780 eAsnProAspPheAlaLysTyrIleAsnAlaGlyAsnAsnLeuValGln 1796
2746 .....GGCAAAACCGGACAGNGTGCAGA 2768
1797 SerMetSerValPheGlySerAsnThrAlaIleThrGlyLysValAs 1813
2769 CACGCGCGCGCGCGCTGCGCGCTCCCTATTAATCCGTTACAGC 2814
1813 pAlaAsnThrGlnSerValGlnGlnProLeuLeu...ValThrPro 1827

seq_name: SwissProt_40:OMP_RICRI
seq_documentation_block:
ID OMP_RICRI STANDARD; PRT; 1654 AA.
AC Q53047;
```

DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas5) (rOmpB)
 DE (rOmp B) [contains: 120 kDa surface-exposed protein (Surface protein
 antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
 GN OMPB.
 OS Rickettsia rickettsii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiense; Rickettsia.
 OX NCBI_TaxID=783;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R;
 RX MEDLINE=92167802; PubMed=1724278;
 RA Gilmore R.D. Jr., Cleple W. Jr., Policastro P.F., Hackstadt T.;
 RT "The 120 kDa outer membrane protein (rOmp B) of Rickettsia
 rickettsii is encoded by an unusually long open reading frame:
 RT evidence for protein processing from a large precursor";
 RL Mol. Microbiol. 5:2361-2370(1991).
 RN [2]
 RP SEQUENCE OF 279-1654 FROM N.A.
 RC STRAIN=R;
 RX MEDLINE=90136087; PubMed=2515418;
 RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
 RT "Cloning, expression and sequence analysis of the gene encoding the
 RT 120 kD surface-exposed protein of Rickettsia rickettsii";
 RL Mol. Microbiol. 3:1579-1586(1989).
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIOLENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 CC LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X16353; CA34403.1;
 DR InterPro: IPR003858; rOmpA_rOmpB.
 DR Pfam: PF02708; rOmpA_rOmpB; 1.
 KW Antigen; S-layer; Cell wall.
 FT CHAIN 1 1333 120 KDA SURFACE-EXPOSED PROTEIN.
 FT DOMAIN 1181 1654 32 KDA BETA PEPTIDE.
 FT DOMAIN 1181 1188 POLY-TMR.
 SO SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;

 alignment_scores:
 Quality: 198.50 Length: 1660
 Ratio: 0.289 Gaps: 69
 Percent Similarity: 41.446 Percent Identity: 17.831
 alignment_block:
 US-09-303-518D-651 x OMPB_RICRI ..
 Align seg 1/1 to: OMPB_RICRI from: 1 to: 1654
 217 GTNTACAAACAAAAAGGAGTTGGTCGCAATCATCATGACAAAGCCCC 266
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 185 VALPheasn.....LeuAlaAsnProThrInGlnLysAlaIsp 197
 267 GATGATGATTTTCTGTGTGTCTCCGCGTAAGAGGCTGGCGGCGATTGGTGG 316
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 197 Oleuille.....LeuG 201
 317 GCGATCAATATATGTGTGAGCGTGACATTAAGCGGCGCTATATAACAGCTT 366

201 |||||.....: ||||| ||||| |||||
 201 LyspsmAlaValIleAlaAsnGlyValAsnGlyThrLeuAsnValThr 217
 367 GATTTTGGTCGCGAAGGAAGNAATCCGATCGACGCGTTTCTTACCA 416
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 218 AsnGlyPheIleGlnValSerAsnLys.....SerPheAl 229
 417 AATTGGAAGAAAGAAATATATTAGCTTGACATTCACACCTTACAGC 466
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 229 aThrValLysAlaIleAsnIleAlaAspGlyGlnGlyIleIlePheAsn. 245
 467 GCGATTANCAATATGCGCGTTTGCATTAATTGTTCACGATGCAAGACT 516
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 246ThrAspAlaAsn 250
 517 GTGCAATAGCAGATGACATGAGGGGAATACCTTATCCGATTAAGAAA 566
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 251 AlaAsnThrLeuAsnLeuGlnAlaGlyThr..... 261
 567 ATATCCGAGCGTGTCCGATGCGCTCAGACACCATATTGGCGTATG 616
 261 261
 617 ATGATGACAAACACGCGATTATCTACTCGGC.....GCA 654
 262ThrIleAsnPheThrGlyThrAspGlyThrGly 272
 655 TGGTTAATGGCGCAATACATATGCGAGGTTGGGAATTAATGGCGT 704
 ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 273 ArgLeuValLeuLeuSerLysHisAlaIleAlaThrAsnPheAsnIleTh 289
 705 AATTGTTGAGGCGCATGTGCGCGCATGCCAGACGATGACCGCTATGC 754
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 289 GylSerLeuGlyGlyAsnLeuLysGlyValIleGlnPheAsnThrValA 306
 755 CGATTGACAGTGGCGGCGAGCGAGCGGTGCCAATGTTATTATGAC 804
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 306 lAlaLysGly.....Gln 310
 805 AAACAAACAATTAATGGCTGCTCAGCGAGTTTACAAACGCGCTACC 854
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 311 LeuThrIleAsnAlaGlyAlaIleAlaAsnAlaValIleGlyThr 324
 855 TTATTCCGCGAGGAAACGTTTCCAGCTGATACGCAAGATTGTTCT 904
 324 324
 905 ACGATGACATTACAGAGCGCATACATACCGTCTTTTGAACCGCGC 954
 324 324
 955 AGTACGAGCATTTTCTTACATCCACACAGCGTACGGGT..... 999
 325AsnAsnGlyAlaGlyArgAl 331
 1000ACGTTACAGAAACCAACGAAAGGTTTCCCAATCCCAAGC 1039
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 331 aAlaGlyPheValValSerValAspAsnGlyLysValAlaThrIleAspG 348
 1040 TTAAGTACAGACAGTCCGACTGTTTGACGAATCTTGAATGAACATG 1089
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 348 LysGlnValTyrAlaLysAspMetValIleGlnSerAlaAsnAlaThrGly 364
 1090 AAACACACAGTT.....TACCGGACGAGGCGTGTATCA 1124
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 365 GlnValAsnPheArgHisIleValAspValGlyValAspGlyThrAla 381
 1125 GTACCGTCCA.....AGTTAAACAACGCGTGAACACTT 1159
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 381 aPheLysThrAlaLaserLysValThrIleThrGlnAspSerAsnPheG 398
 1160 CTTTATCGATTACGGCAAC.....GCCAAA 1185
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```
398 LysAsnThrAspPheGlyAsnLeuAlaIleGlyValProAsnAla 414
1186 CTCATCTTATCAACACATC.....AACCAAGC.... 1215
    ::::::::::::::::::::|||
415 IleThrLeuThrGlyAsnPheThrGlyAspAlaSerAsnProGlyAsnTh 431
1216 .GGGGCGGCTTGTATTTGAAAGTGATTTTACGCTCGCTCGCTGAAACA 1264
    ||||| ::::::::::::::::::::|||
431 rAlaGlyValIleThrPheAspAlaAsnGlyThrLeuGluSerAlaSerA 448
1265 ACGAAACGTGGCAAGCGCGGCTTCATATCAGTGAAGACATGACGTT 1314
    ::::::::::::::::::::|||
448 laAsp.....AlaAsnValAlaValThrAsnAsnIleThrAla 460
1315 ACTTGAAGATTAACGGCGTGGCAACGACCGCTGCCAAATCGGCA 1364
    ::::::::::::::::::::|||
461 Ile..GluAlaSerGlyAlaGlyValValGlnLeuSer..... 472
1365 AGGACGCTGCACGCTTCAGCCAAAGGGAAACCAAGGCTGCATC...A 1411
    ||||| ::::::::::::::::::::|||
473 .GlyThrHisAlaIleGluLeuArgLeuGlyAsnAlaGlySerIlePhe 489
1412 CGGTGGCGGACGCTACAGTCTTTGGATCGACGACGACGATAAAGC 1461
    ::::::::::::::::::::|||
489 ySLeuAlaSpGlyThrValIle.....AsnGly 498
1462 AAAAACAAGCCTTAGTGAATCGGCTTGNTACGCGGACGAGGTAGCGT 1511
    ||| ::::::::::::::::::::|||
499 LysValAlaGlnThrAlaLeuValGlyGlyAlaLeuAlaIleGlyThrII 515
1512 GCAACTGAATGCCGATTAATCAGTTCACCCCGAC..... 1545
    ::::::::::::::::::::|||
515 eThrLeuAspGlySerAlaThrIleThrGlyAspIleGlyAsnAlaGlyG 532
1545 ..... 1545
532 lYAlaAlaAlaLeuGlnArgIleThrLeuAlaAsnAspAlaLysLysThr 548
1546 .....AAACTATTTCCGCTTTCGCGGCGGACGCTTGA 1580
    ::||| ::|||
549 LeuThrLeuGlyGlyAlaAsnIleThrGlyAlaGlyGlyThrIleAs 565
1581 TTTA.....AAGGGCATTCGCTTTCGTCACCGCTATTCAAAT... 1620
    ||||| ::::::::::::::::::::|||
565 pLeuGlnAlaAsnGlyGlyThrIleLysLeuThrSerThrGlnAsnAsnI 582
1621 .....ACCGATGAAGGGCGATGAT 1641
    ||||| ::::::::::::::::::::|||
582 leValValAspPheAspLeuAlaIleAlaThrAspGlnThrGlyValVal 598
1642 GNCNATCATATGCGACACAACATCACCGCTTACCATTTACAGGATGA 1691
    ::::::::::::::::::::|||
599 AspAlaSerSerLeuThrAsnAlaGlnThrLeuThrIleAsnGlyLysII 615
1692 AAGTATTACACAACGAGGTGAAGATATCATATGACTTAATTC... 1737
    ::::::::::::::::::::|||
615 eGlyThrIleGlyAlaAsnAsnLysThrLeuGlyGlnPheAsnIleGlyS 632
1738 ..ACGAAGAATAATGCTTACAAACGGTTGGTTGGCAGAGAAGAT... 1779
    ||||| ::::::::::::::::::::|||
632 eSerLysThrValLeuSerAsnGlyAsnValAlaIleAsnGlnLeuVal 648
1779 ..... 1779
649 IleGlyAsnAspGlyAlaValGlnPheAlaHisAspThrTyLeuIleTh 665
1780 ...ACGACCAAGACAGCGGCGCTCACTGTTTACACAGCGCGCGG 1825
    ||||| ::::::::::::::::::::|||
665 rArgThrThrAsnAlaIleGlyGlnGlyLysIleIlePheAsnProVal 682
1826 CAGAAGACCGCACCCNGCTGCTTCCGGGAGCAAAATTTAAACGCG... 1872
    ::::::::::::::::::::|||
682 alAsnAsnGlyThrThr...LeuAlaIleGlyThrAsnLeuGlySerAla 697
1872 ..... 1872
698 ThrAsnProLeuAlaGluIleAsnPheGlySerLysGlyValAsnValAs 714
1873 .....ANACTCA 1879
714 pThrValLeuAsnValGlyGlyGlyValAsnLeuTyralThrAsnIleT 731
1880 CGCAACCAAC.....GGCAAACTGTTTCAGCGCGGACGACCGCA 1920
    ||| ::::::::::::::::::::|||
731 hrThrThrAspAlaAsnValGlySerPheValPheAsnAlaGlyGlyThr 747
1921 CCGACACCTTACATCATTTAGCAACGCGGTGTCMAAATGAAGAT 1970
    ::::::::::::::::::::|||
748 AsnIleValSerGlyThrValGlyGly.....GlnGlnGlyAs 760
1971 CCCACAGAGAAATCGTGTGGACACGACGTGATCNAACCC... 2013
    ::::::::::::::::::::|||
760 nLysPheAsnThrValAlaLeuGlnAsnGlyThrThrValLysPheLeuG 777
2014 .....ACGTTTAA.....GCGGAAATTTCCATATT 2040
    ||||| ::|||
777 LysAsnAlaThrPheAsnGlyAsnThrThrIleAlaIleAsnSerThrLeu 793
2041 CAGGCGGCGGACGCGGTGATTTCCGCAATGTTGCCAAATGGAAGCGCA 2090
    ||| ::::::::::::::::::::|||
794 GlnIleGlyGlyAsnTyThrAlaAspCysValAlaSerAlaAspGlyTh 810
2091 T..... 2091
810 rGlyIleValGluPheValAsnThrGlyProIleThrValThrLeuAsnII 827
2091 ..... 2091
827 ySglnAlaIleProValAsnAlaLeuLysGlnIleThrValSerGlyPro 843
2092 .....TGNCATTTGACGACATCAGCCCAAC 2117
    ::::::::::::::::::::|||
844 GlyAsnValValIleAsnGlnIleGlyAsnAlaGlyAsnHisIleGlyAl 860
2118 AGTTTTGGTGGCGCACCCGATCAAGCCATACATC..... 2154
    ||||| ::::::::::::::::::::|||
860 aValThrAspThrIleAlaPheGluAsnSerSerLeuGlyAlaValValP 877
2155 .....TGT 2157
877 heLeuProArgGlyIleProPheAsnAspAlaGlyAsnThrMetProLeu 893
2158 ACAAGTTCGAGCGACGNGGCTGCACAATTTGT..... 2190
    ||| ::::::::::::::::::::|||
894 ThrIleLysSerThrValGlyAsnLysThrAlaLysGlyPheAspValPr 910
2191 .....GTGAAANAANCATTACGACGATGAAGTGA 2221
    ||||| ::::::::::::::::::::|||
910 oSerValValLeuGlyValAspSerValIleAlaAspGlyGlnValI 927
2222 TTGCTTATTGACACACGACGACGACGACGACGACGACGACGACGAC 2271
    ||| ::::::::::::::::::::|||
927 leGly.....AspGlnAsnAsnIleValGlyLeuGlyLeu 938
2272 NACGNTNNTTMAANCTC..... 2289
    ::::::::::::::::::::|||
939 GlySerAspAsnGlyIleIleValAsnAlaThrThrLeuTyralGlyII 955
2290 .....NCGGCGNTGCGNCACTMAAGCAATCTTA 2320
    ||||| ::::::::::::::::::::|||
955 eSerThrLeuAsnAsnAsnGlnGlyThrValThrLeuSerGlyValP 972
2321 GTGCAAAATGCGGATACGTTAT..... 2343
    ::|||
972 roAsnThrProGlyThrValTyrcIleLeuGlyThrGlyIleGlyAlaSer 988
```



```
2344 .....AC 2345
989 LysPheIysGlnValIThrPheThrThrAspTyrAsnAsnLeuGlyAsnI 1005
2346 AGTCAGCCCAACGCGCCCAAAAGCGACCTTAGCCTGTG.....G 2389
1005 eIleAlaThrAsnAlaThrIleAsnAspIleValIThrValIThrThrGlyG 1022
2390 GCAATGCCCAACACATTTAATCAAGCCACATTAACGGCGCAACACATCG 2439
1022 LylIleAlaIleIleGlyPheAspGlyLysIleThrLeuGlySerValAsn 1038
2440 GNTTCGGGCATGCTCTCATTTAAT.....CTAAGCAACACGCGCG 2480
1039 GlyAsnGlyAsnValAlaArgPheAlaAspGlyIleLeuSerAsnSerThrSe 1055
2481 ACAAAACGCGAGCTGTAGCGCTTCCGACACAGCTTAAGCAACAGTAAGCC 2530
1055 rMetIleGlyThr...ThrLysAlaAsnAsnGlyThrValIThrTyrLeuG 1071
2531 ATTCCGCACTCAAGGCAATGTCCTACCGCATAGGCAATATTCAT 2580
1071 LysAlaIlePheValGlyAsnIleGlyAspSerAspThrProVal..... 1085
2581 TTTGAAACAGCGCGCTTACCGGACACTCAAGCGCGACAGCAAGANACAGC 2630
1086 ...AlaSerValArgPheThrGlySerAspSerGlyAlaGlyLeuGlnG 1101
2631 A.....TTACACTTAA 2641
1101 yAsnIleTyrSerGlnValIleAspPheGlyThrTyrAsnLeuGlyIleV 1118
2642 AAGACAGCAATGAGACGCTGCGCGTACAGCAGCAATTAAGCAATTAAC 2691
1118 aIAsnSerAsnIleIleLeuGlyGlyGlyThrAlaIleAsnGlyLys 1134
2692 CTTGACAAACGCGCACCTTACACTCAATTCGCGTATCCGACAGATGTGC 2741
1135 IleAspLeuValIThrAsnThrLeuThrPhe.....AlaSe 1146
2742 AGCGCGCAACCGGCAAGNGTGTACAGACAGCGCGCGCGCTTGGCGCC 2791
1146 rGlyThrSerThrTyrGlyAsnAsn..... 1154
2792 GTTCCCTATTATCGTTACACCGCACTCGGTAGATCCGTTTCAAC 2841
1155 .....ThrSerIleGluThr..... 1159
2842 ACGTACGCGTA...AACGCAAAATTGAAC.....NGTCAAG 2876
1160 ThrLeuThrLeuAlaAsnGlyAsnIleGlyHisIleValIleLeuGlnG 1176
2877 AACATTCCGCTTATGTGGAACTCTGGGCTACCGAAGCAACAATTGA 2926
1176 yAla.....G 1178
2927 AGCTGGCGAAAGTTCGAAGNACTTACACTTGGCGGTCAACAATACC 2976
1178 InValAsnThrThrThrThrGlyThrThrThrIleIleValAlaGlnAspAsn 1194
2977 GCGAAGCAACCGGTAGCCTGTGATCA...TTGACGGTAGTGAAGG... 3021
1195 AlaAsnAlaAsnPheSerGlyThrGlnThrTyrThrLeuIleGlnGly 1211
3022 ....AAGACACAACAACGCTG...TCGAAACCTTAATTTCACCCGCG 3064
1211 yAlaArgPheAsnGlyThrLeuGlySerProAsnPheAlaValIThrGlyS 1228
3065 AAAAGCAACAGCTGTAGTCCGCGCGGTGTACCAATCATCCGCAAA 3114
1228 eTAsnArgPheValAsn.....TyrSerLeuIleArgAla 1239
3115 GACGGCGAG.....TTCCGCTGCAATAATCCGGT 3143
1240 AlaAsnGlnAspTyrValIleThrArgThrAsnAsnIleGlnAsnValIa 1256
3144 CAAGAACAACAGCTTTCGACAAACTCGGCAGAGCAAGCAACCAAAAAAC 3193
1256 LThrAsnAspIleAlaAsnSerProPheGlyGlyAlaProGlyAlaAsp. 1272
3194 AGCGGAAAAAGACACAGCCCAAGCCTTGACGCGCTGATGGCGCGCG 3243
1273 .....GlnAsnValIThrThrPheValAsnAlaIleThr 1282
3244 CGCGATCGCGCGCAAAAGACAGAAAGCTTGCCGAA...CCGCGCGCGCN 3290
1283 AsnThrAlaIleAlaTyrAsnAsnLeuLeuAlaIleLysAsnSerAlaAsnSe 1299
3291 GCGAGCGCGGCAAAATGTGCGCATTTATGACGCGGAGAGAAAGAAAAAC 3340
1299 rAlaAsnPheValGlyAlaIleValIThrAspThrSerAlaAlaIleThrA 1316
3341 GGTGTCAAGCGCGATTAAGACAGCGCMTTGGCGAAACAGCCGAGCGGAA 3390
1316 snValGlnLeuAsp.....LeuAlaLysAspIleGlnAlaGln 1328
3391 ACCGCGCGGNTACACCGCTTCCCGCGCGCGCGCGCGCGGGA 3440
1329 .....LeuGlyAsnArgLeuGlyAlaLeuArgTy 1338
3441 TTTGCGCGCAACCGCAGCGCCACCGCACTCAACCCCAACCGCAGCGCG 3490
1338 rLeuGlyThrProGlnThrAlaGlnMetAlaGlyProGlu..... 1351
3491 ACCTGATNAGCGTTATGCCAATAGCGGTTGACTGATTTCCGCGCAG 3540
1352 .....AlaGlyAla 1354
3541 CTCACAGCGTTTTCGCGGTACAGAGCAA...TTGACCGCGCTGTTCG 3587
1355 IleSerAlaIleAlaIleAlaGlyAspGlyAlaIleLysAsnVal..... 1369
3588 CGAAGACCGCGCAAGCGCTTGGACAGCAGCAGCATCCGAGNACACCAAC 3637
1370 .....AlaTyrGlyIleThrPalalysProPheTyrThrAspAlaH 1383
3638 ACTAC...CGTTCGCAAGATTTCCGCGCTACCGGCAACAAACGCACTG 3684
1383 IsGlnSerLysGlyGlyLeuAlaGlyTyrLysAlaLysThrThrGly 1399
3685 CGCCAAATCGGTATGCAGAAAAACCTCGCAGCGGCGCGTCCGACATCT 3734
1400 ValValIleGlyLeu..... 1404
3735 GTTTTCGACACACCGGACCGAAACANCTTGCAGACGCGCATCGCACT 3784
1405 .....AspThrLeuAlaAsnA 1410
3785 CGGCAGCGTTTCCGCAAGCGCGCGCTTTCGGCAATACGCGATCGGAGG 3834
1410 spAsnLeuMetIleGlyAlaIleAlaIle.....GlyIleThrLys 1422
3835 TTTGCAATCGGCATACGACAGCGGCGCGGTTTACGACGCGCANTCTNC 3884
1423 ThrAspIle.....LysHisGlnAspTyrLysGlyLysLysThr 1436
3885 AGAGCGCATGAGAGCAAAATCCGCGCGCGCGTGTGATTCAGGATTC 3934
1436 rAspValAsnGlyPheSerPheSerLeu.....TyrGlyAlaG 1449
3935 AGCGACGA.....TACCGCGCGGTTTCGGCGGATTCGGCGATC 3972
1449 GlnGlnLeuValLysAsnPhePheAlaGlnGlySerAlaIlePheSerLeu 1465
3973 GAACGCTACATCGGC...GCAACGCGCTATTTCGTC..... 4005
```



```
991  ergInserLeuGluSerGluAspAlaIrgThrSerCysLysAspArgIu 1007
      :::::|||||:
1418  GCACGCGTACAGTCATTTGGATCGACGACGACGATAAAGCAAAAAA 1467
      |||:|||||:
1008  AlaValAlaIglInAla.....ProArgH 1015
      :
1468  CAAGCTTAGTGAATCGGCTGN..... 1492
      : |||:|||||:
1015  sAlaProValValaIrgThrProSerIleGlnProSerLeuLeuProHIsA 1032
      :
1493  .....TCACGCGACGGGTA 1507
      |||:|||||:
1032  laAlaProPheAlaLysSerHisLeuValHisGlySerSerProGlyVal 1048
      :
1508  CGGTGCACTGAATCGCGATATCACTTCACCCGCAAACTATTTTC 1557
      |||:|||||:
1049  .....MetGlyThrSerValAlaThrSerAlaSerLysIleI 1061
      :
1558  GGCTTTCGGCGGAGCGTTGGATTTAAACGGGCAATCGCTTTCGTCCA 1607
      :|||:|||||:
1061  eProGlnGlyAlaAspSer.....ThiMetLeuAlaThrLysThrV 1075
      :
1608  CGCTATTCAAATACCGATGAAGGGCGATGTGNCNATTCATATGCCA 1657
      :
1075  aLlysHisGlyAlaProSerProSerHis.....ProIleSerAlaPro 1089
      :
1658  CAACACATCCACCGTTACCATTTACAGGATGAAGAAAGTTACACACCG 1707
      |||:|||||:
1090  GlnGlnLeuAlaAlaAlaLeuAlaIrgGlnMetAlaSerGlnAlaPr 1106
      :
1708  AGTGTAGATATATCAATAGACTTATTTACAGCAAAAGAAATTCCTACA 1757
      |||:|||||:
1106  oAlaValAsn...ThrLeuThrGluSerThrLeuLysAsnValaProGlnV 1122
      :
1758  CGGTGCTTGGCGAGAAAGATACGACCAAAAGACGGCGGCTCAACC 1807
      :|||:|||||:
1122  aValAsnValaGlnGluLeuLysAsnAsnProAlaThrProSerThrAla 1138
      :
1808  TTGTTTACCAGCCCGCGAGAAAGCCGACCCNGCTGTTTCGGCGGA 1857
      :|||:|||||:
1139  MetGlySerSerValProLysSerThrAlaLysThrProHisProVala 1155
      :
1858  ACAAAATTTAAGCGCACATCACGCAAAACAAACGCTGTTTTCAG 1907
      :|||:|||||:
1155  uThrProValAlaAlaAsnGlnAlaLysGlnLysLeuIleAsnSerL 1172
      :
1908  CGGACAGCGACACCGCGCTACATCATTTAGGAAGCGGTGTCAA 1957
      :|||:|||||:
1172  euLysProSerGlyProThrPro.....AlaSerGlyGln 1183
      :
1958  AATGGAAAGTATCCACAGAGAAATCGTGTGGACACAGCATGTGATC 2007
      ||| |||
1184  .....LeuSerSerGlyAspLysAlaSerG 1192
      :
2008  NACCCGACGTTTAAAGCGGAAATTTCCATATTCAGGGCGGCGACGG 2057
      :|||:|||||:
1192  yThrAlaLysIleGlnThrAlaValThr.SerThrProSerAlaSerGly 1208
      :
2058  GATTTCCGCAATGTTGCCAAAGTGGAAAGCGATTCATTTGAGCAATC 2107
      :|||:|||||:
1209  GlnPhe.....Se 1211
      :
2108  ACGCCCAAGCACTTTTGGTGTGACCGCATCAAGCCATACATCTGT 2157
      :|||:|||||:
1211  rLysProPheSerPhe.....SerProSerGlyThrGlyP 1223
      :
2158  ACAAGTTTGGAGTGCACNCGTGCACAATGTGTGCAGAAAANCAATTAC 2207
      :|||:|||||:
1223  HeAsnPhcGlyIle..... 1227
      :
2208  CGACGATAAAGTATGCTTCATTCATAGACNCGACNTNAGCGGCANTG 2257
      ||| ||| ||| :|||:|||||:
1228  .....Ile.ThrProThrProSerSerAsnPhet 1237
      :
2258  TNAGCNNTNCCNATNACGNTNNTTAAANCTCNCGGGCNTGNCACTN 2307
      :
1237  hr..... 1237
      :
2308  AANGCAATCTTAGTGCAAATGCGGATACGCTTATACAGTACGCCAA 2357
      :|||:|||||:
1238  .....AlaIleGlnGlyAlaThrProSerThr.....LysG 1248
      :
2358  CGCCGCCAAACGCAACCTTAGCCTCGTGGGCAATGCCAACACACAT 2407
      :|||:|||||:
1248  uSerSerGlnProAspAlaPheSerSerGlyGlyGlySerLysProSer 1265
      :
2408  TTAAT.....CAAGCCACATTTAAAGCGCAAC 2433
      :|||:|||||:
1265  yGluAlaIleProGluuSerSerProProSerGlyIleThrSerAlaSer 1281
      :
2434  NCATCGGNTTGGCGCAATGCTTCATTTATCTAAGCAACAAGCCGCACA 2483
      :|||:|||||:
1282  AsnThrThrProGlyGluProAlaIleSerSerSerAlaProValAlaPr 1298
      :
2484  AAAGCGCAGTCTGACGCTTCCGACAAAGCTAAGCAAGCAAGCCATT 2533
      :|||:|||||:
1298  oSerGlyThrAlaLeuSerThrThrSerSerLysLeuGlnThrProPro 1315
      :
2534  CCGCAGTCAACGCGCAATGTC.....TCCTAAGCCGATTAAGCAGTA 2574
      :|||:|||||:
1315  eLrLysLeuGlyGluLeuLeuPheProSerSerLeuAlaGlyGluThrLeu 1331
      :
2575  TTCATTTTGAACAGCGCTTACCGGACACCTGACGGCGAGC..... 2619
      :|||:|||||:
1332  GlySerPheSerGlyLeuAlaG...ValGlyGlnAlaAspSerThrLy 1347
      :
2620  .....AAGANACAGCATTTACCTTAAAGACAGCGAATGAGCGC 2659
      :|||:|||||:
1347  sProThrAsnLysAlaSerSerThrSerLeuThrSerThrGlnProThrL 1364
      :
2660  TGCGGTACAGCGAGGAATTAGCAATTTAAACCTTGCAACGCCACCAT 2709
      :|||:|||||:
1364  yStrThrSerGlyValProSerGly...PheAsn.PheThrLapProProVa 1379
      :
2710  A.....CACTCAATTCGCGCTATC.....GCCAGATGCTGC 2741
      :|||:|||||:
1379  lleGlnLysHisThrGluProProValThrSerSerAlaThrThrHis 1396
      :
2742  AGCGCGCAACCGCGCAGNGTGTACAGACCGCGCGCGCGCTTCGGCGC 2791
      :|||:|||||:
1396  eValAlaIleProProAlaAlaThrSerThr.SerSer..... 1407
      :
2792  GTTCCCTATTTATCCGTTACACCGCCACCTCGGTAGAAATCCGTTTCA 2841
      :|||:|||||:
1408  .....ThrAlaValaPheGlySerLeuProValaThrS 1418
      :
2842  ACGGTGACGGTAAACGCAATTTGACNCTGACAAATTCGCTTAT 2891
      :|||:|||||:
1418  eAlaIleGlySer.SerGlyValIleSerPheGlyGlyThrSerLeuSerAl 1434
      :
2892  GTGCGAACTC.....TTGCGCTACGAGACGCAAAATTAAGCTGG 2932
      :|||:|||||:
1434  aGlyLysThrSerPheSerPheGly.SerGlnGlnThrAsn..... 1447
      :
2933  CGGAAAGTTCGGAAGNACTTACACCTTGGCGGCAACATACCGGCAC 2982
      :|||:|||||:
1448  .....SethrValProProSe 1453
      :
2983  GAACCCGTAAAGCTCGATCAATTGACGGTAGTGAAGGGAAGAACAA 3032
      :|||:|||||:
1453  rAlaPro..... 1455
      :
3033  ACCGCTTCGGAACCTTATTTACACCTTCGCAAAAGCAACAGCTGATG 3082
      :|||:|||||:
1456  ....ProProThrThrAlaAlaThrProLeuProThrSerPheProThr 1470
```

```

3083 CCGGCGCGGTGATTACCACTCCGCAAGAGCGGAGTTCGCGCTG 3132
      :::: ||| ||||| ||| ||| |||
1471 LeuSerPheGlySerLeuSerSerAlaThrThrProSerLeu.Prom 1487
      :::: ||| ||| ||| ||| |||
3133 CATAAATCGGTCMAAGAACAGAG.....CTTTCGCA 3164
      :::: ||| ||| ||| ||| |||
1487 eSerAlaGlyArgSerThr.GluGluAlaThrSerSerAlaLeuProGlu 1503
      :::: ||| ||| ||| ||| |||
3165 CAATCTCGGCAAGCAGAAAGCCAAAACAGCG.....GAA 3202
      :::: ||| ||| ||| ||| |||
1503 uLysProGlyAspSerGluValSerAlaSerAlaSerLeuLeuGluG 1520
      :::: ||| ||| ||| ||| |||
3203 AAGCAACGCGCAAGCCTTGACGGCTGATTGCGCGCGCGCGATGCC 3252
      :::: ||| ||| ||| ||| |||
1520 LuGluGlnSerAlaGlnLeuProGlnAlaProProGlnThrSerAspSer 1536
      :::: ||| ||| ||| ||| |||
3253 GCGCAAAAGACAGAAAGCGTTGCCGACCGCGCGCGCGCGCGGGA 3302
      :::: ||| ||| ||| ||| |||
1537 ValLysLysGluProValLeuAlaGlnProAlaValSerAsnSergly.. 1552
      :::: ||| ||| ||| ||| |||
3303 AATGTGCGCATTTATGACGGCGAGAGAGAAAAACGGGTGCAGCGG 3352
      :::: ||| ||| ||| ||| |||
1553 .....ThrAlaAla 1556
3353 ATAAAGACAGCGCTTTGGGAAACAGCGCAAGCAACCGCGCGGNT 3402
      :::: ||| ||| ||| ||| |||
1556 eSerThrSerLeuValAla.....LeuSerAlaGluAlaThrProAla 1570
      :::: ||| ||| ||| ||| |||
3403 ACCACGCGCTTCCCGCGCGCGCGCGCG.....CCGCGGGATTTGC 3445
      :::: ||| ||| ||| ||| |||
1571 ThrThrGlyValProAspAlaThrGlnAlaValProAlaSerSer 1587
      :::: ||| ||| ||| ||| |||
3446 C.....GCAACCG 3453
1587 rPheSerValProGlyGlnThrAlaValThrAlaAlaIleSerSerA 1604
      :::: ||| ||| ||| ||| |||
3454 CAGCCCAACCGCA.....ACCTCAACCCCAACCGCAGCGGACT 3494
      :::: ||| ||| ||| ||| |||
1604 IglYProValAlaValAlaGlnThrSerSerThrProIleAlaSer.SerTh 1620
      :::: ||| ||| ||| ||| |||
3495 GATNAGCGCTTATGCC..... 3510
      :::: ||| ||| ||| ||| |||
1620 rThrSerIleValAlaProGlyProSerAlaGluAlaAlaIlePheGlyT 1637
      :::: ||| ||| ||| ||| |||
3511 .....AATAGCGTTGAGTGAATT.....TCCGCCACGCTC 3543
      :::: ||| ||| ||| ||| |||
1637 hValAlaThrSerGlySerSerValPheAlaGlnProProAlaIleSer 1653
      :::: ||| ||| ||| ||| |||
3544 AACAGCGTTTCCGCGTACAGACGAATTCAGCGGTTTCCGGA 3593
      :::: ||| ||| ||| ||| |||
1654 SerSerAlaPhe.AsnGlnLeuThrAsnAsnThrAlaThrAlaPro...S 1669
      :::: ||| ||| ||| ||| |||
3594 CCGCCGCAACGCGTTTGGACAGCAGCATCCGAGACCAACACTAC 3643
      :::: ||| ||| ||| ||| |||
1669 eRAlaThrProValPheGlyGlnValAlaAlaSerThrAlaProSerLeu 1685
      :::: ||| ||| ||| ||| |||
3644 GTTCGAAGATTTCGCGCGCTACCGCAACACGACCTGCGCAATC 3693
      :::: ||| ||| ||| ||| |||
1686 PheGlyGlnGlnThrGlySerThrAlaSerThrAlaAlaIleThrProGlu 1702
      :::: ||| ||| ||| ||| |||
3694 GGTATGCAAGAAACCTCGGACGCGCGCGCTGCGCATCTGTTTCCGA 3743
      :::: ||| ||| ||| ||| |||
1702 nVal.....SerSerSerglyPheSerSerProAlaPheGlyT 1715
      :::: ||| ||| ||| ||| |||
3744 CAGCGGACG..... 3754
      :::: ||| ||| ||| ||| |||
1715 hThrAlaThrProGlyValPheGlyGlnThrThnPheGlyGlnAlaSerVal 1731
      :::: ||| ||| ||| ||| |||
3755 .....AAACANCTTCAGACGAGCATGCGCA.....ACTGCGCA 3789
      :::: ||| ||| ||| ||| |||
1732 PheGlyGlnSerAlaSerSerAlaAlaSerValPheSerPheSerGlnPr 1748
      :::: ||| ||| ||| ||| |||

```

```

3790 CCGCTTGCCACGCGCGCGTTTTCGGGCAATACGCAATCGCAGTTGCA 3839
      :::: ||| ||| ||| ||| |||
1748 oGlyPheSerSerValProAlaPheGlyGlnProAlaSerSerThrProT 1765
      :::: ||| ||| ||| ||| |||
3840 CATCGGCATC.....AGCACGGGC 3858
      :::: ||| ||| ||| ||| |||
1765 hThrThrSerGlySerValPheGlyAlaAlaSerSerThrSerSer 1781
      :::: ||| ||| ||| ||| |||
3859 GCGGTTTATGACGCGCATCTTTCAGACAGCATGAGCAAAATC... 3906
      :::: ||| ||| ||| ||| |||
1782 SerSerPheSerPheGlyGlnSerSerProAsnThrGlyGlyLeuTh 1798
      :::: ||| ||| ||| ||| |||
3907 ....CGCCCGCGGTGCTGATTCAGCATTCAGCAGCATACCGCGCG 3952
      :::: ||| ||| ||| ||| |||
1798 eGlyGlnSerAsnAlaProAlaPheGlyGlnSerProGlyPheGlyGln 1815
      :::: ||| ||| ||| ||| |||
3953 GT.....TTCGCGGATTGCGCATGGAACCGTACATCGCGCGCACG 3993
      :::: ||| ||| ||| ||| |||
1815 IyGlySerValPheGlyGlyThrSerAlaAlaThrThrThrAlaAlaThr 1831
      :::: ||| ||| ||| ||| |||
3994 CGC.....TATTTGCTCCAAAAGCGATTAACCGCTACGAAACGTCAA 4037
      :::: ||| ||| ||| ||| |||
1832 SerGlyPheSerPheGlyGlnAlaSerGlyPheGlySerSerAsnThrGl 1848
      :::: ||| ||| ||| ||| |||
4038 T.....ATCGCACCCCGGTCTTGGCTTCAACCGT 4069
      :::: ||| ||| ||| ||| |||
1848 ySerValPheGlyGlnAlaAlaSerThrGlyGlyIleValPheGlyGln 1865
      :::: ||| ||| ||| ||| |||
4070 ACCGNGC..... 4077
      :::: ||| ||| ||| ||| |||
1865 InsSerSerSerSerglySerValPheGlySerGlyAsnThrGlyArg 1881
      :::: ||| ||| ||| ||| |||
4078 .....GCAATTAAGCAGATTAATTCATTCAAACCGCGCAACAC...AT 4118
      :::: ||| ||| ||| ||| |||
1882 GlyGlyGlyPhePheSerSerGlyLeuGlyGlyLysProSerGlnAspAla 1898
      :::: ||| ||| ||| ||| |||
4119 NTCCATCANCCTTATTTNAGCTGTCTTACCATGCGCGCTCGCGCA 4168
      :::: ||| ||| ||| ||| |||
1898 AsnLysAsnProPheSerSerAlaSer.....GlyGlyPheGly 1912
      :::: ||| ||| ||| ||| |||
4169 AAGTCGCAACGCGTCAATACGCGNGTATG.....GCT 4203
      :::: ||| ||| ||| ||| |||
1912 eRThrAlaThrSer...AsnThrSerAsnLeuPheGlyAsnSerGlyAla 1927
      :::: ||| ||| ||| ||| |||
4204 CAGCATTTTCGCAAAACCCGAGTGGGAAATGGGCTAAACCGCAAT 4253
      :::: ||| ||| ||| ||| |||
1928 LysThrPheGlyGlyPheAlaSerSerPheGlyGlnGlnLysProTh 1944
      :::: ||| ||| ||| ||| |||
4254 CAAGGTTTACGCTGTCCNTCCACGCTGCCGCCCAAGGCCCAAC 4303
      :::: ||| ||| ||| ||| |||
1944 rGlyThrPheSerSerGlyGlyGlySerValAlaSerGlnGlyPheGlyP 1961
      :::: ||| ||| ||| ||| |||
4304 TGAAGCGCAACACAGCGCGGC 4326
      :::: ||| ||| ||| ||| |||
1961 hSerSerProAsnLysThrGly 1968
      :::: ||| ||| ||| ||| |||
seq_name: SwissProt_40:FRPC_NEIMB
seq_documentation_block:
ID FRPC_NEIMB STANDARD: PRT: 1829 AA.
AC O9JYV5:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Iron-regulated protein frpc.
GN FRPC OR NMB1415.
OS Neisseria meningitidis (serogroup B).
OC Bacteria, Proteobacteria, beta subdivision, Neisseriaceae, Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
MEDLINE=20175755; PubMed=10710307;

```

[illegible]

1082 yAsnAspThrLeuIleGlyAlaGlyAsnAspTyrLeuGluGly 1099
779 GCGGTGCGCAATGTTATTATGACAAACAAACAAATTAATGCGTCGTC 828
1099 ergLySerAspThrTyrValPheGlySerGlyPheGlyGln 1112
829 AACGAGTTTACAAACCGCGTACCCTTAT 1113
1113AspThrValTyrAsnTyrAspTyrAlaThrGlyAlaGly 1125
870 AACGCTTTCAGCTGATCGCAAGATGTTGTTACGATGACAT 915
1125 SASP 1125
916TACAGCGCATACACATACCGTCTNTTGGAA 948
1136 laAspMetLeuThrPheThrArgGluGlyAsnHisLeuIleLeuAla 1152
949 CCGCGCATACGCGACATTTTCC 992
1153 LysAspGlySerGlyGlnValThrValGlnTyrTyrPheGlnAsnAspGly 1169
993 TACGGGTACGGGTACAGAAACCAACGAAAGTTCACATCCAAAG...C 1039
1169 ySerGlyAlaTyrArgLeuAspGluLeuHisPheAspAsnGlyValLeu 1186
1040 TTAAGTACAGACAGTCCGA...CTGTTGACGAATCTTTGAATGAAC 1086
1186 euAspValAlaThrValLysGluLeuValGlnGlnSerThrAspGlySer 1202
1087 GATAAAGAACAGTTTACGCGGCGGTGTTATCATGACGCGCCAAAG 1136
1203 AspArgLeuTyrAlaTyrGlnSerGlyAsnThr 1213
1137 GTTAAC 1153
1214 LeuAsnGlyGlyLeuGlyAspAspTyrLeuTyrGlyAlaAspGlyAsp 1230
1154 ACCCTTCTTTTATGATTACGACGCAACCATCTTATTCACAAAC 1203
1230 sPLeuLeuAsnGlyAspAlaGlyAsnAspSerIleTyrSerGlyAsnGly 1246
1203 1203
1247 AsnAspThrLeuAspGlyGlyGluGlyAsnAspAlaLeuTyrGlyTyrAs 1263
1204ATCACCAAGCGCGCGCGTGTGATTGTAAGT 1239
1263 nGlyAsnAspAlaLeuAsnGlyGlyGluGlyAsnAspHisLeuAsnGlyG 1280
1240GATTTTACGCTCGCGCGCAACACACGAACTGGCAA 1278
1280 laAspGlyAsnAspThrLeuIleGlyGlyAlaGlyAsnAspTyrLeuGln 1296
1279 GCGCGC 1304
1297 GlyGlySerGlySerAspThrTyrValPheGlyLysGlyPheGlyGlnAs 1313
1305 CAGTACCGTTACTGGAA 1323
1313 pAlaValTyrAsnTyrAspTyrAlaThrGlyAlaGlyAspIleIleArgP 1330
1324GTAAACGCGTGGCAACGACGCGCTG 1350
1330 heThrAspGlyIleThrAlaAspMetLeuThrPheThrArgGluGlyAsn 1346
1351TCCAAATCGGCAACGACGCTGACGTTCAAGC 1385
1347 HisLeuLeuIleLysAlaLysAspGlySerGlyGlnValThrValGlnSe 1363
1386 CAAGGCGCAACGACGCTGATGACGCTGGGCGACGATACAGCATTT 1435
1363 rTyrPheGlnAsnAspGlySer 1375
1436 TGATACGACGACGACGATTAACGCAAAACAAAGCCTTAGTGAATC 1485
1375 leAspIleLeuHisPheAspAsnGlyLys 1386
1486 GCGTGTGTCAGCGCGGAGGAGTACGAGTGAATGCGATTAATCAGTT 1535
1387 AspValAlaThrValLysGluLeuValGlnGlnSerThrAspGlySer 1402
1536 CAACCCGCAACACTGATTTGCGCTTTCGCGGAGCGATTTGA 1585
1403AspArgLeuTyr...AlaTyrGlnSerGlyAsn...ThrLeu 1415
1586 ACGGCGATTCGCTTTCGCGGAGGAGTTCACAAATTCAGTAAAGGCGG 1635
1415 smGly...GlyLeuGlyAspAspTyrLeuTyrGlyAlaAsp...GlyAsp 1429
1636 ATGATTGNCNATCATTAATCCACACACATCCACGTTACATTAACG 1685
1430 AspLeuLeuAsnGlyAspAlaGlyAsnAspSerIleTyrSerGlyAsnGly 1446
1686 GAATGAAGATATTACACAAACGAGTGTAAAGATATCATATGACTTAAT 1735
1446 yAsnAspThrLeuAsnGlyGlyGlyGly...AsnAspAlaLeuTyrGlyT 1462
1736 ACAGCAAGAAATTCCTACACGCTTGGCGGAGAAAGATACGACC 1785
1462 yAsnGlyAsnAspAlaLeuAsnGlyGlyGly 1473
1786 AAAACGACGCGCGCTCAACCTTGTTCACGCGCGGACGAGACG 1835
1474AsnHisLeuAsn...GlyGluAspGly 1482
1836 CACCCNCTGCTTTCGCGGAGCAAAATTTA 1867
1482 yAsnAspThrLeuIleGlyGlyAlaGlyAsnAspTyrLeuGluGlyGly 1499
1868 ACGGCAACATCAGCAACAAACGCAACCTTTTTCAGCGGACGACG 1917
1499 ergLySerAspThrTyrValPheGlyLysGlyPhe...GlyGlnAsp 1513
1918 ACACCGACCGCTCAACATCATTTAGAGAGCGGTGTCAAAATATGGAAG 1967
1514 AlaValTyrAsnTyrAspTyr 1521
1968 TATCCCAAGAGAAATCGTGTGGGACACGATGATCCGCG 2013
1521 aThrGlyArgLysAspIleIleArgPheThrAspGlyIleThrAlaAspM 1538
2014ACGTTTAAACGCAAAATTTCCATATTCAGGCGGCGGCGGTG 2058
1538 eLeuThrPheThrArgGluGlyAsnHisLeuLeu 1549
2059 ATTTCCCGCAATGTTGCCAAAGTGAGGAGGATTTGACGAATCA 2108
1550IleLysAlaLysAspGlySerGlyGlnValThr 1560
2109 CGCCCAAGCAGTTTGTGTGTCGACCGCATCAAGCCATACATCTGTA 2158
1561 ValGlnSerTyrPhe 1565
2159 CAGCTTGGAGTGCAGCGCTGACAAATGTGTGCAANAANACATTAAC 2208
1566GlnAsnAspGlySerGlyAlaTyrArgLeuAspGluLeuHisPheAsp 1581
2209 GACGATTAAGTATGCTTCATTGACT 2237
1582 AsnGlyLysValLeuAspValAlaThrValLysGluLeuValGlnGlnSe 1598
2238 GACGACATNAGCGGCAANTGATNAGCTNNCCNATNAGCTNNNTTNAANC 2287
1598 rThrAspGlySerAspArgLeuTyrAlaTyrGlnSerGlySerThrLeuA 1615

2626 ACAGCATTTACCTTAAGAGACGCAATGAGACGCGCCGCTGAGGACGCA 2675
571 HisValSerProArgLysArgIleu..... ArgI 581
2676 ATTAGCAATTTAAACCTTTGACAAACGCCACCATTAACATCAATCCGCT 2725
581 upHeGluLysValSerLeuGlnAsp..... AsnIlycysV 593
2726 ATGCGCAGATCTGCAAGCGCAACCGCAGNGTGTCAACACGCGC 2775
593 aAlasmsnIlySerGIyAlaSerSerGIyAlaIagIyLgIy.... 607
2776 CCGCGCGGCTGCGCGCTTCCCTATTATCCGTACCGCCCAACTGGCT 2825
608 ...LysArgSerArgAlaLysGlyThrSerThrSerProAlaGlyL 623
2826 AGAATCCGCTTCAACACAGCTGACGGTTAAAGCGCAAAATTGAACGTCAAG 2875
623 sAlaSerPrometAsnLeuAlaProProGlnIlyLysPro..SerProser 639
2876 GAACATTCGCTTATATGTCGAACCTTTCGGCTACCGAAGGACAAATTG 2925
640 ProGlySerSerSerSerSerThrSerProAlaThrLeuSerTrngIn..P 656
2926 AAGCTGCGGGAAGTTCCGAAGNACTTACAC.....CTTGGGGGTCA 2969
656 roThrArgLeuAsnSerSerLysSerIleHisSerLeuGlnGlyLysEr 672
2970 CAATACCGGCAACGAACCCGTAAAGCTGCATCAATTGACGGTATCGAGAG 3019
673SerGIySerGIySe 677
3020 GGAAGACACAACAACCGCTGTCGAAACCTTAATTGACCCGTGCAAAC 3069
677 rSerSerSerSerSerSerGIyLysGIyCysGIyAspHisProAla... 692
3070 GAACACGTGATGCGCGCGCGGTGAGTTACCACTCACTCCGCAAAAGACG 3119
693AlaIleIleSerAsnValHisHisProGlnHisSer 704
3120 CGAGTTCGCTCGCATTAATCCGCTGCAAGAACAGAGCTTTCGCACAAC 3169
705 MetTyrGlnProser...SerSerSerLysProArgAlaLeuLeuThrSe 720
3170 TCGGCAA.....GCGAGAACGC 3186
720 rProLysSerProAspValSerGIySerAsnGIyLysGIyLysSerP 737
3187 AAAAAACAGCGGGAATAACACAACCGCGCAAGCCCTTGACGGGCT... 3230
737 roSerHisThrGlyThrLysLysArgSerProProLysSerAlaGlySer 753
3231GATTGCGCGCGCGCGGATGCGCGCAAAAGACAGAAACGCTTG 3274
754 ProValAspTyrGIyHisSerPheTyrArgAsPro..... 765
3275 CGGAACCGCGCGCGGCGAGCGCGGGAATAATGTGCGCATTTATGACGCG 3324
766TyrAlaGlyA 769
3325 GAGGAGAGAAAAAAG...GCTGCAAGCGGATTAAGACAGCGCGTTTGGC 3371
769 IaGIyArgProSerThrSerGIySer..AlaSerGlnAsp.....LeuSe 783
3372 GAACACGCGCAAGCGGAAACCGCGCGGNTACCAACCGCCTTCCCGCGCG 3421
783 rProProArgSerSerProAlaSerProAlaThrThr..ProArgThrVal 799
3422 CCGCGGCGCGCGCGCGGATTTGCGGCAAC...GCACCCGCAACCGCA 3468
800 ProLysLysThrAlaSerIleArgTyr..GluPheAlaSerProSerAla 816

```

3469  COTCAACCCCA.....ACCGAGCGGACGATCGAT 3497
      :::::
816  eTserSerGysPrroSerProGlyAspArgSerAlaSerProProGlu 832
      :::::
3498  NAGCGGTGTGGCCAAATAGCGGTTTGAGTGAATTTTCCGCGACGCTCAACA 3547
      Arg.....ArgHisMetGlnG 838
      :::::
3548  GCGT.....TTTCCCGGTACAGAGCAATTTGACGCGCGTGTGGCGAA 3591
      :::::
838  nGlnProHisLeuGlnArgSerSerProLeuHisTyrTyrMetGlyPro 855
      :::::
3592  GACCGCGGCAACGCGNCTTTGGAAACGNCATCCGCGAACGCAACGCTA 3641
      :::::
855  roProProGln..ValAsnGlyAsnGlySerAlaGlySerProThrSer 870
      :::::
3642  CGGTTGCAAGATTTTCCGCGCTACCGGCGCCACAAACCGACCTGGCGCAA 3691
      871.....AlaProProThrSerAlaSerAlaAlaAla... 881
      :::::
3692  TCGGTATGCAGAAAAACSTCGAGCGGCGCGCTCGGATCTGTTTTCG 3741
      882.....ValAlaAlaAlaAlaAlaAlaAlaAlaAlaTyr 892
      :::::
3742  CACACCGGACCGCAAAACACTTCG.....ACAGCGCATCGCGAA 3782
      :::::
892  rLleProSerProSerLleTyrAsnProGlyLleSerThrLeuAlaAla 909
      :::::
3783  CTGCGGACCGGC..... 3793
      :::::
909  euArGlnHisnProLeuTTrMetHisTyrGlnThrGlyAlaSerPro 925
      :::::
3794  ..TTGCCACGCGCGCGCTTTTCGCGCAATACGCGATCGGAGTTTCGAC 3840
      :::::
926  LeuLeuSerProHisProGlnProGlnGlyGlySecAlaAlaAlaAlaAla 942
      :::::
3841  ATCGGCATACAGCAGCGGCGCGGCTTTTACGACGCG.....GCANTCTNTC 3884
      :::::
942  aaAlaAlaAlaAlaAlaTyrLeuSerProGlnSerAlaTyrHisAlaPheAla 959
      :::::
3885  AGACGCGCATCGCGGAAATCGCGCGCGCGCTGCATTTACGCGCATTC 3934
      :::::
959  yTrAsnGlyValGlyAlaAlaValAlaAlaAlaAlaAlaAlaAlaPhe 975
      :::::
3934  ..... 3934
      :::::
976  GluGlnProAlaProSerProHisThrHisProHisLeuAlaHisProHis 992
      :::::
3935  .....AGGACAGATACGCGCGCGG 3953
      :::::
992  sGlnHisProHisProAlaAlaLeuThrThrHisHisSerProAlaHisT 1009
      :::::
3954  TTTCGGCGGATTCGGATCGAACCGTAAACATCGGCGGACGCGCATTTTCG 4003
      :::::
1009  euAlaThrProGlyLeuThrAspSerSerThrAspGlnMetSerAlaThr 1025
      :::::
4004  TCCAAACACGCGATTACCGGTACGAAACGCTCATATGCGCACCCCGGT 4053
      :::::
1026  SerSerHisAlaArgThrAlaSerThrSerProSerSerSerGlnAlaSerAl 1042
      :::::
4054  CTTCGCTCAACCGNATACCGNCGCGGCGCATTTAAGGACG...ATTATTCAT 4099
      :::::
1042  aSerSerSerAlaAlaThrSerGlyAlaSerSerSerAlaMetPheHisT 1059
      :::::
4100  ..TCAACCGCGGCGACACATATTCATCAACNCTT 4132
      :::::
1059  hTserSerLeuAlaGlyAsnGluGlnSerSerAspLeu 1070
      :::::
seq_name: SwissProt_40:PM20_CHLPPN
seq_documentation_block:
ID      PM20_CHLPPN      STANDARD:      PRT:  1723  AA.
NC      Q9Z812; Q9K2C1; Q9R359;

```


617 hrserAsnLysAlaAlaaspLcLgLYgLYLeuTYrCYsGLyAsp 633
1243 TTTACGGCTCG.....CCTGAACACACGA 1268
634 ValThrLeuThrAsnLeuThrGLyLysThrLeuPheGlnGluAsnSerSe 650
1269 AACGTGGCAGCGCGCGCTTCATATCAGTAGACAGATACCTTACTT 1318
650 rGLuLysHisGLyGLyLeuSerLeuAlaSerGLyLysSerLeuThr 667
1319 GG.....AAAGTAAC 1329
667 ethrSerLeuGLuSerPheCYsLeuAsnAlaAsnThrAlaLysGLuAsn 683
1330 GGC..... 1332
684 GLyGLyGLyAlaAsnValProGLuAsnLeValLeuThrPheThrTYrTh 700
1332 1332
700 rProThrProAsnGLuProAlaProValGlnGlnProValTYrGLyLc 717
1333GTGGCAACGACCGCCTGTCCAAATCGCCAAAGCAGCGCTGCAC 1377
717 lAleuValThrGLyAsnThrAlaThrLysSerGLyGLyLleTYrThr 733
1378 GTTCAGCCAAAGGGGAAACCAAGGCTCGATC..... 1410
734 LysAsnAlaAlaPheSerAsnLeuSerValThrPheAspGlnAsnTh 750
1411 .AGCGTGGCGCAGCGTACAGTCAATTTGGATCAACGAGCAGACAGCAATAAG 1459
750 rSerSerLcLysnGLyGLyAlaLeuLeuThrGlnLysAlaAlaAspLys. 766
1460 GCAAAAAACAGCCTTAGTGAATCGGCTGTGNTCAGCGGAGGGGTACG 1509
767 ..ThAspCYsSerPheThrTYrLle..... 774
1510 GTGCACATGATGCCGATATCATGTTCAACCCGACAACTCTATTTCGG 1559
775 ThrAsnValAsnLleThrAsn.....AsnThrAlaThrGL 786
1560 CTTTCGGCGCGAGCGTTGGATTTAAAGCGGCACTTCCTTCCTCCACC 1609
786 YAsnGLyGLyLys.....LleAlaGLyGLyLysAlaHisPheAsp 800
1610 GTATTCAAAAT.....ACCGATGAAGGGCGCATG 1638
800 rGlnLeAspAsnLeuThrValGlnSerAsnGlnAlaLysLysGLyGLy 816
1639 ATTGNCNATCATTAATGCCACAAACAAATCCACGTTACCATTAAGGG.. 1686
817 ValTYrLeuGlnAspAlaLeuLleLeuGlnLysVal...LleThrGLySe 832
1687AATGAAGATTACACAAACGAGTGTAGATATCATTAAGACTTA 1732
832 rValSerGlnAsnThrAlaThrGlnSerGLyGLyLle..... 845
1733 ATTACAGCAAGAATTT.....GCCATCAACGCTGGTTGGCGAG 1773
846 ..TYrAlaLysAspLleGlnLeuGlnAlaLeuProGLySerPheThrLle 861
1774 AAAGAT.....ACGACAAAGAGAGCGGCGGCT 1802
862 ThrAspAsnLysValGlnThrSerLeuThrThrSerThrAsn..... 875
1803 CAACCTGTTTAAACAGCCCGCGCAGAGAGACGCCACCCGCTGCTTTCG 1852
876LeuTYrGLyG 879
1853 GCGGACAATTTAAAGCAACATCAACGCAAAACAAACGCAACTGTTT 1902
879 lYgLYlLleTYrSerSerGLyAlaValThrLeuThrAsn..... 891

1903 TTCAGCGCAGACCGACACCGACCGCTACATTCATTTAGAAGCGGTG 1952
892 lLeSerGLy..... 894
1953 GTCAAAAAATGGAAGTATCCCAACAGAGAAATCGTGGACAAACGACT 2002
895ThrPheGLyLleThrGLyAsnSerValLle..... 904
2003 GGATCAACCGCACGTTTAAAGCGGAAATTTCCATATTCAGGCGGG... 2049
905AsnThrAlaThrSerGlnAspAlaAspLleGlnGLyGLy 918
2050 ...CAGCGGTGATTTCCCGCAATGTGCCAAAGTGAAGGCGGATTGNA 2096
919 lLeTYrAlaThrThrSerLeuSerLleAsnGlnCYsAsnThrProLleLe 935
2097 TTTGACCAATCAACGCCACAGCTGTTTGTGTCGACCGCATCAAGCC 2146
935 uphSerAsnAsnSerAlaAla..... 942
2147 ATACAAATCTGTACAGCTTCGACCTGCACNGCTGCACAAATTTGTGAA 2196
943ThrLysThrSerThrThLys 950
2197 NAANCAATTACGACGATAGATAGATGCTTCATTCATTAACAGACACMT 2246
951 GlnlAlaGLyGLyAlaAlaPheSerAlaAlaValThrLcLcLcLc 967
2247 NAGCGGCANTGTNAGCNCTNCCNATNACGNTNNTNAAANTCNCNGGC 2296
967 nserGlnProLlelLePheLeuAsnAsnSerAlaLysSerGlnAlaThrt 984
2297 NTGCNNCACTNAANGCAATCTTAGTCAAAATGGCATACAGTTATACA 2346
984 hAlaAlaThrAlaGLyLysLysAspSerCYsGLy..... 995
2347 GTACGACACACGCGCACCCCAACGCAACCTTACGCTGGCGCAATGC 2396
996GLyAlaAlaAlaAlaAsnSerValThrLeuThrAsnAsnTr 1009
2397 CCAAGCAATTTAATCAAGCCACATTAAACGCAACNCATCGGNTCCG 2446
1009 ogLlLleThrPhe.....LysGlnAsnTYrAlaGLuThG 1021
2447 GCAATGCT.....TCATTTAATCTAAGCAACACGCC..... 2478
1021 lYgLYlAlaLleGLyCYsLleAspLeuThrAsnGLySerProProAlaGLys 1037
2479GCACAAACGCGAGTGTGACGCTTCGCAACAGCATAGGC 2519
1038 ValSerLleAlaAspAsnGLySerValLeuPheGlnAspAsnSerAlaLe 1054
2520 AAC.....GTNAGCCATTCG 2536
1054 uAsnATrGLyGLyAlaLleTYrGLyGLuThrLleAspLleSerArgThG 1071
2537 CACTCAAC.....GCCAT..... 2550
1071 lYAlaThrPheLleGLyAsnSerSerLysAsnAspLYserAlaLleCYs 1087
2551GTTCCTTAGCGCATAGGCGATTCCTCATTTTGAATA 2588
1088 CYsSerThrAlaLeuThrLeuAlaProAsnSerGlnLeuLlePheGln 1104
2589 CAGCGCTTACCGGACAACTAGCGGACGAGCAAGAMACAGATTACACT 2638
1104 nAsnLysValThrGLuThrThrAlaThrThLysAlaSerLleAsnAsnL 1121
2639 TA.....AAGACAGCAATGACGCTGCGG 2664
1121 euGLyAlaAlaLleTYrGLyAsnAsnGLuThrSerAspLleThrLleSer 1137

```

2665 TCAGCAGCGAATTAGCAATTAAACCTTGACACGCCACCATTTACACT 2714
      :::::  |||  :::::  ::  |||
1138 LeuSerIaGluAsnGlySerIlePhePheLysAsnAsnLeuGlyThrAl 1154
      :::::  |||  :::::  ::  |||
2715 CAATTCGCCCTATCGCCACGATGCTGCGAGCGCGCAACCGCGAGNGTGT 2764
      :::::  |||  :::::  ::  |||
1154 aThrAsnLysIYrCySserIleAla.....GlyAsnValL 1166
      ::  |||  ::  |||  ::  |||  ::  |||
2765 GAGACAGCGCGCGCGCGCTTCGCCCTATATATACGTTACACCG 2814
      ::  |||  ::  |||  ::  |||  ::  |||
1166 ySpherThrAlaIleGluAlaSerAlaGlyLysAlaIleSerPheTyrrsp 1182
      ::  |||  ::  |||  ::  |||  ::  |||
2815 CCAACTTCGGTA.....GAATCCCGTTTCAACAGCTGACGCTAA 2855
      :::::  |||  :::::  ::  |||  :::::  |||
1183 AlaValaSnValSerThrLysGluThrAsnIaGlnGluLeuLysLeuAs 1199
      ::  |||  ::  |||  ::  |||  ::  |||
2856 CGGCAATTTGAACNGTCAAGACACTTCGCTTATGTGCGAAGCTCTTC. 2904
      |||  :::::  |||  |||  ::  |||  |||
1199 nGluLysAlaThrSerThrGlyThrIleLeuPheSerGlyLueuHisc 1216
      ::  |||  ::  |||  ::  |||  ::  |||
2905 .....GGCTACCGAAGCGCAATG..... 2925
      ::  |||  ::  |||  ::  |||  ::  |||
1216 LuAsnLysSerTyrlleProGlnLysValIThrPheAlaHiscGlyAsnLeu 1232
      ::  |||  ::  |||  ::  |||  ::  |||
2926 .....AAGCTGCGCGAAAGTTCGAGAGNACTTACACC..... 2958
      |||  |||  |||  |||  |||  |||
1233 IleLeuGlyLysAsnIaGluLeuSerValIserPheThrGlnSerPr 1249
      ::  |||  ::  |||  ::  |||  ::  |||
2958 ..... 2958
1249 oGlyThrThrIleThrMetGlyProGlySerValLeuSerAsnHiscL 1266
      ::  |||  ::  |||  ::  |||  ::  |||
2959 .....TTGGGGGTCAACACATACCGCGCAACACCGCTAAGC 2994
      ::  |||  ::  |||  ::  |||  ::  |||
1266 ySGLuIaGlyGlyIleAlaIleAsn.....ValIle 1277
      ::  |||  ::  |||  ::  |||  ::  |||
2995 CTCGATCAATTGACGTAGTAGAGGAAAGACAAC 3030
      ::  |||  ::  |||  ::  |||  ::  |||
1278 IleAspPheSerGluIleValProThrLysAspAsn 1289
      ::  |||  ::  |||  ::  |||  ::  |||
seq_name: SwissProt_40:OMPb_RICJA
seq_documentation_block:
ID OMPb_RICJA STANDARD; PRT; 1656 AA.
AC 006653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas5) (OMPb)
DE (Comp B) [contains: 120 kDa surface-exposed protein (Surface protein
antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide].
GN OMPb
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-YH:
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein omp B of Rickettsia
japonica.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
SIMILARITY).
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
(BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
LAYER WITH HEXAGONAL SYMMETRY.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPb FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AB003681; BAA20138.1; -
DR InterPro: IPR003858; rompA_rompB.
DR Pfam: PF02708; rompA_rompB; 1.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1338 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.
FT DOMAIN 528 533 POLY-GLY.
FT SEQUENCE 1656 AA; 168097 MW; 3132A69CD5999F CRC64;

```

alignment_scores:

```

Quality: 191.00      Length: 1512
Ratio: 0.284          Gaps: 73
Percent Similarity: 44.511 Percent Identity: 19.246

```

alignment_block:

```

US-09-303-518D-651 x OMPb_RICJA ..

```

```

Align seg 1/1 to: OMPb_RICJA from: 1 to: 1656

```

```

208 GATATTGAGGTGTACACAAAAAGGAGTGTGCGCAATCAATGAC 257
      |||  :::::  |||  :::::  ::  |||  ::
347 AspGlyGlnValIYrAlaLys.....AspMetValIleGlnSerAlaAs 361
      ::  |||  ::  |||  ::  |||  ::  |||
258 AAAAGCCCGCATGATGATTTCT.....GTGGTGGCGCTAACG 298
      ::  |||  ::  |||  ::  |||  ::  |||
361 nAlaAsnGlyGlnValaSnPheArgHiscIleValaIspValGlyIleAspG 378
      ::  |||  ::  |||  ::  |||  ::  |||
299 GCGTGGCGGCGATGCGGCGATCAATATATGTTGACGTCGACATAC 348
      ::  |||  ::  |||  ::  |||  ::  |||
378 LyrThrThrAlaPheLysThrAlaIaIaSerIleValaIaIleThrGlnAsn 394
      ::  |||  ::  |||  ::  |||  ::  |||
349 GCGGCGCTATACACAGCTGATTTGCT.....GCGGAGGAGGANA 389
      ::  |||  ::  |||  ::  |||  ::  |||
395 SerAsnPheGlyThrThrAspPheGlyAsnLeuAlaIaGlnValIYrVa 411
      ::  |||  ::  |||  ::  |||  ::  |||
390 TCCGATCAGCACCGCTTTTCTTACCAATGTGAAAGAAATATATATA 439
      |||  |||  |||  |||  |||  |||
411 lProAsp.....ThreThrIleuThrGlyAsnPhet 422
      ::  |||  ::  |||  ::  |||  ::  |||
440 AGCCTGACAAATTCACACCCCTTACACGCGCATTCATCATATGCGCGTTTG 489
      |||  :::::  |||  :::::  ::  |||  ::
422 hrgLyspAlaAsnAsnProGlyAsn.....Thr 431
      ::  |||  ::  |||  ::  |||  ::  |||
490 CATAAATTTGTACAGATGCGAAGCCTGTCGAATGACGACATGAC 539
      ::  |||  ::  |||  ::  |||  ::  |||
432 AlaGlyValIleThrPheAlaIaAsnGlyThrLeuAlaSerAlaSer 448
      ::  |||  ::  |||  ::  |||  ::  |||
540 GGGGAATACCTATTCGATTAAGAAATAATACCGGCGGTGCGCATCG 589
      ::  |||  ::  |||  ::  |||  ::  |||
448 lAspAlaAsnValAlaValThrAsnAsnIleThrAlaIleGluAlaSer 464
      ::  |||  ::  |||  ::  |||  ::  |||
590 GCTCAGACACACCACTATGCGGCTTATGATGATG.....ACAACAGCGC 633
      ::  |||  ::  |||  ::  |||  ::  |||
465 Gly.....ValGlyValaIaGlnLeuSerGlyThrIsthrAl 477
      ::  |||  ::  |||  ::  |||  ::  |||
634 GAT.....TTATCTACTCCGCGCA...TGCTTAATTTGGCGGCAATA 673
      |||  :::::  |||  :::::  ::  |||  ::
477 acIuLeuArgGlyAsnAlaGlySerValPheLysLeuAlaAspGlyT 494
      ::  |||  ::  |||  ::  |||  ::  |||
674 CACATATGACGGGTGGGAAATATGCGGTATTTAGTTGACGCGCAT 723
      ||  :::::  |||  |||  ::  |||  |||
494 hrValIleAsnGlyLysValaSnGlnThrVal.....LeuValGlyGly 508

```



```

4006 .....CAAAAGCGGATTACCGCTACGAAACGTCA 4036
1481 aaanglyAsnMetSerLySglInleAlaIaIaGlyThrMetAsnIleThrAspMet 1498
4037 ATATGCCCAACCCCGGCTTGGCTTACACGNTACCGNCGGCGCATTAAG 4086
1498 hr.....PheIlyGlyAsnLeuThr 1504
4087 GCAGATTATTCATTCAACCGCGCAACATTCATCCATCCACCTTATT 4136
1505 ValGlyTyrAspTyrAsnIaMetGlnGlyValLeuValThrProMetAl 1521
4137 MAGCCTGCTCTATACCGATCCGCT..... 4161
1521 aglyLeuSerTyrLeuLysSerSerAspIuAsnTyrLysGluThrGlyT 1538
4162 .....TCGGCAAGTCCGCAACCGCTCAAT 4188
1538 hrThValAlaAsnLysGlnValAsnSerLysPheSerAspArgThrAsp 1554
4189 ACCGCGTATGTGCTCAGGATTTCGGCAAAACCGCGCAGT...GGCGAATG 4235
1555 leuIleValGlyAlaLysValAlaGlyThrMetAsnIleThrAspPh 1571
4236 GGGCGTAAAGCCGCAATCAAGTTTCACGCTG 4269
1571 eAlaValTyrProGluValHisAlaPheValAla 1582

```

seq_name: SwissProt_40:OMP_RICPR

seq_documentation_block:

```

ID OMPB_RICPR STANDARD; PRT; 1643 AA.
AC Q53020; Q9ZCM0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rOMPb)
DE antigen) (110 kDa outer membrane protein ompb); 32 kDa beta peptide).
GN OMPB OR SPAP OR SPA OR RP704.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_taxid=782;
RX SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BREINL;
RT MEDLINE=91045972; PubMed=2122457;
RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;
RT "Characterization of the gene encoding the protective paracrystalline-
RT surface-layer protein of Rickettsia prowazekii: presence of a
RT truncated identical homolog in Rickettsia typhi."
RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BREINL;
RA Moron C.G., Yu X.J., Walker D.H.;
RT "Sequence analysis of ompb of Rickettsia prowazekii."
RL Submitted (JUN-1999) to the EMBL/Genbank/DBD databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-MADRID E;
RX MEDLINE=9039499; PubMed=9823893;
RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
RA Scharitz-Ponten T., Alsmark U.C.M., Podowski R.M., Neeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).
RN [4]
RP PARTIAL SEQUENCE.
RC STRAIN-BREINL;

```

```

RX MEDLINE=92114896; PubMed=1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNBR fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii."
RL Mol. Immunol. 29:95-105(1992).
RN [5]
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE=92104668; PubMed=1729180;
RA Hackstadt T., Messer R., Cieplik W. Jr., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kDa outer
RT membrane protein of rickettsiae: identification of an avirulent mutant
RT deficient in processing."
RL Infect. Immun. 60:159-165(1992).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPb FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M37647; AAA26390.1; ALT_INT.
DR EMBL; AF161079; MAD42234.1; -.
DR EMBL; A0235273; CAAL5140.1; -.
DR InterPro: IPR003858; rOMPb_rOMPb.
DR Pfam; PF02708; rOMPb_rOMPb; 1.
KW Antigen; S-layer; Cell wall; Complete proteome.
FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.
FT VARIANT 257 257 32 KDA BETA PEPTIDE.
FT VARIANT 1010 1010 Y -> A (IN STRAIN BREINL).
FT VARIANT 1450 1450 A -> D (IN STRAIN BREINL).
FT VARIANT 178 179 AA -> VC (IN STRAIN BREINL).
FT CONFLICT 191 201 TTQEARPLFGA -> INRSSTHYLVS (IN REF. 1).
FT CONFLICT 212 212 T -> I (IN REF. 1).
FT CONFLICT 313 313 Q -> L (IN REF. 1).
FT CONFLICT 1104 1104 D -> G (IN REF. 2).
FT CONFLICT 1123 1123 T -> S (IN REF. 2).
SQ SEQUENCE 1643 AA; 169854 MW; 735FD392E6346CC CRC64;

```

alignment_scores: Quality: 190.50 Length: 1376
 Ratio: 0.303 Gaps: 66
 Percent Similarity: 45.712 Percent Identity: 18.750

alignment_block:

US-09-303-518d-651 x OMPB_RICPR

Align seg 1/1 to: OMPB_RICPR from: 1 to: 1643

```

661 ATATGCCGCAATACA.....CATATGCCGCTGGGCAAT.. 696
257 ValGlyGlyAsnThrIleAsnPhaGlyIleAspGlyThrGlyLysIle 273
697 .....ATATGCCGCTAATTA 709
273 uValleuValSerLysAsnGlyAlaAlaThrGluPhaAsnValThrGlyT 290
710 GTTTCAGCGCGCATGTCGCCCATGCCACGACTATAGCCGCGGATT 759
290 hrLeuGlyLysAsnLeuLysGlyIleIleGluLeuAsnThrAlaAlaVal 306
760 GCAGTTCGCGCAGCGACAGCGGTTCCGCAATGTTATTATTGACAAAC 809

```

[illegible][illegible]


```

2397 CCAAGCAACATTTAATCAAGCCACATTAAGCGCAAC..... 2433
      |||
      :|||
820 PProlethr.....ValthrLeuasnLysGlnGlyAlaTyrPheG 834
2434 .....NCATCGGNTTCGGCAATGCTTCATTTAAT 2463
      |||
      :|||
834 lYValleuLysGlnVallelleleSerGlyProGlyasnleValPheasn 850
2464 .....CTAAGCAAC..... 2472
      |||
      :|||
851 GluileGlyasnValGlyleValHISGlylleAlaAlaasnSerlleSe 867
2473 .....AAGCCGCAAAAACCGCAGTCTGACGCTTTCGACAAACGCT 2514
      |||
      :|||
867 rPheGluasnAlaSerleuGlyThrSerleuPheleuProSerGlyThrP 884
2515 .....AAGCAACGTAAGCCATTCGCGACATCAAC 2544
      |||
      :|||
884 rOleuaspValleuthrThrleLysSerThrValGlyasnGlyThrValasp 900
2545 GGCATGTCCTCCCTACCGCATAGCGATATTCATTCATTTGAAAACAGCCG 2594
      |||
      :|||
901 ...AsnPhasnAlaProleValValValSerGlyleAspSerMetil 916
2595 CTTTACCGGACAACTCAGCGGAGCAAGANACAGCATTAACCTTA.... 2640
      |||
      :|||
916 eAsnasnGlyGlnlelleGlyAspPylsYasnlelleleAlaLeuSerl 933
2641 .....AAAGACAGCAATGAGCGCTCCCTCAGCGACGAATTAGCAAT 2685
      |||
      :|||
933 euGlySeraspasnSerleThrValasnAlaasnThrleuTyrSerGly 949
2686 TTAACCTT.....GACAACGCGACCATTAACATTCGCGCCCA 2726
      |||
      :|||
950 lLeArgThrThrLysasnasnGlnGlyValThrPheThrThrThrThr 966
2727 TCGCCACGATGCTGACGCGCGCAACCGCGCAGNGTGCAGACACGCCG 2776
      |||
      :|||
966 tProasnasn.....ProGlyThrleThrleThrleGlyleGlyl 978
2777 GCGCGCTTCGCGCGCTTCCTATTATCGGTTACACCGCAACTTCGCGTA 2826
      |||
      :|||
978 euGlnasnGlySerProLysleuLysGlnValThrPheThrThrThrThr 994
2827 GAATCCCGT.....TTCAACACGCTGACGCGTAAC..... 2856
      |||
      :|||
995 AsnasnLeuGlySerllelleAlaasnasnValThrleasnAspTyrVa 1011
2856 ..... 2856
1011 lThrleuthrThrGlyGlylleAlaGlyThrAspPheasnAlaLysileT 1028
2857 .....GGCAAAATTGAACNGTCAAGAAACATTCGCTTATGTCGGAATC 2901
      |||
      :|||
1028 hrleuGlySerValasnGlyasnAlaasnValAlaArgPheValAspSerThr 1044
2902 TCGCGCTACCGAAGCAGACAAATTGAAGTGGCGGAAGTTCGGAAGNAC 2951
      |||
      :|||
1045 PheSeraspProArgSerMetlleValAlaThrGlnAlaasnLysGlyTh 1061
2952 TTACAC.....TTGGCGGTCAACATTAACGCG.....AAGC 2993
      |||
      :|||
1061 rValThrThrleuGlyasnAlaAlaLeuValSerasnleGlySerleuasnPT 1078
2984 AACCCGTAGCCTCGATCAATTAAGGTAGTGAA.....GGGAAA 3024
      |||
      :|||
1078 hrProValAlaSerValArpPheThrGlyasnAspSerGlyAlaGlyleu 1094
3025 GACAACAAACCGCTGTCGAAACCTTAATTC.....AC 3059
      |||
      :|||
1095 GlnGlyasnleThrThrSerGlnasnleAspPheGlyThrTyrasnLeuth 1111
3060 CCTGCAAAAGACACGTCGATGCCGCGCGTGGCTTACC..... 3100
      |||
      :|||
1111 rlleuasnSerasnVallleleuGlyGlyGlyThrThrAlaAlaasn 1127
3101 .....AACTCATCCGC 3111
1128 GlyGluileAspLeuLeuthrThrAsnAsnleullePheAlaasnGlyThrSe 1144
3112 AAAGACGGGAGCTTCGCGCTGATATTCGCTCAAAAGAAAGAGCTTC 3161
      |||
      :|||
1144 rThrThrPylaspasnThrSerlleSerThrThrleuasnValSer..Se 1160
3162 CGACAAACTCGGCAAGCAGAGAACCAAAAACAGCGGAA..... 3201
      |||
      :|||
1160 rGlyasnlelleGlylnVallleAlaGluAspAlaGlnValAlaAla 1177
3202 .....AAAGCAACGCGCAAGC... 3219
1177 hrThrThrGlyThrThrThrThrleLyslleGlnaspAsnAlaasnAlaasn 1193
3220 .....CTTGACGCGCTGATTCGCGCGCGCGCGCATGCCG 3254
1194 PheSerGlyThrGlnAlaTyrThrleuThrleGlnGly.....Al 1207
3255 GGAAGACAGCAAGAGCTTCCGCAACCGCGCGCGCGCGCGGAGAA 3304
      |||
      :|||
1207 aArgPheasnGlyThrleuGlyAlaProasnPheAlaValThrGlySerA 1224
3305 ATGTCGCTATTCAGCGGAGGAGAAAGAAACGCGGTGACGCGGAT 3354
      |||
      :|||
1224 snllePheVal.....LysTyrGluLeulleArg...AspSerasn 1236
3355 AAAGACAGCGCTTTCGCGAAACAGCGGAGGAGGAGGAGGAGGAGTAC 3404
      |||
      :|||
1237 GlnaspTyrValleuthrArgThrAsnaspVal...LeuasnValThr 1252
3405 CACCGCTTCGCGCGCGCGCGCGCGCGCGCGGATTTCCGCAACCGC 3454
      |||
      :|||
1252 rThrAlaValGlyasnSerAlaAlaAlaasn..... 1262
3455 AGCCCAACCGCAACCTCAACCCCAACCGGAGCGCATGATNAGCGGT 3504
      |||
      :|||
1263 .....AlaProGlyValSerGlnasnleSerArgCysleuGlySerThr 1277
3505 TATGCCAATAGCGGTTGATGATTTTCGCCACGCTCAGACGCTTT 3554
1278 Asn.....ThrAlaAlaTyrasnSerMetlle 1286
3555 CGCGCTACAGAGCAATTTGAC.....CGCGGTGTTGCCG 3589
      |||
      :|||
1286 uLeuAlaLysaspProSeraspValAlaThrPheValGlyAlaAlaAlaT 1303
3590 AAGACCGCGCAACGCTTTCGAGACAGCGCATCCGCGNACCAAAACAC 3639
      |||
      :|||
1303 hrAspThrSerAlaAlaValThrThrValasnleuasnAspThrGln... 1318
3640 TACCGTTCGCAAGATTTCCGCGCTACCGCCCAACAAACGACTGGGCCA 3689
      |||
      :|||
1319 ...LysThrlnasp..... 1322
3690 AATCGGTATGCAGAAAAACCTCGGAGCGCGCGCTGCGCATCTGTTT 3739
1323 .....leuLeuSerasnArgleuGlyThrleuArgT 1333
3740 CGCACACCGGAGCGAAACANTTTCGACGACGCGCATCGGCAACTGCGCA 3789
      |||
      :|||
1333 yrlLeuSerasnAlaGluThr..... 1339
3790 CGGCTTGCCCGACGCGCGCTTTTCGGCAATACGCGATCGGAGGTTGCA 3839
1339 ..... 1339
3840 CATCGCATTCAGCAGCGCGCGGCTTTTACACGCGCATCTNTCAGACG 3889

```

```

1340 .....SerAspValAlaIleSerAlaThrGlyAlaValaSerSerc 1353
13890 GCATCGGAGCAAAATCCCGCGCGGTGCTGATTCAGCATTCAGCA 3939
1353 LysAspGluAlaGluVal.....SerTyrGlyAlaTrrPala 1364
3940 CGA.....TACCGCGCGGTTCGGCGG 3962
1365 LysProHeTyrAsnIleAlaGluGlnAspLysGlyIleAlaG 1381
3963 ATTC.....GGCATCGAACCGGTAC.... 3981
1381 YTYrLysAlaLysThrThrGlyValValGlyLeuAspThrLeuAla 1398
3982 .....ATCGCGCGCAACCGCTTTCGTC 4005
1398 eAspAsnLeuMetIleGlyAlaAlaIleGlyIleThrLysThrAsp 1414
4006 CAATAACCGGATTACGC.....TACGAACGTCATATCGGCAC 4046
1415 LysHisGlnAspTyrLysGlyAspLysThrAspLeu..... 1428
4047 CCGCGCTTTCGCTTCAACCGTAC 4071
1429 ...GlyLeuSerHeSerLeuTyr 1435

```

seq_name: SwissProt_40:YM96_YEAST

seq_documentation_block:

```

ID   YM96_YEAST          STANDARD;          PRT: 1140 AA.
AC   004893;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 113.1 kDa protein in PRES-FET4 intergenic region.
GN   YMR317W OR YMR924.09.
OS   Saccharomyces cerevisiae (Baker's Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=5288C / AB972;
RA   Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC   -!- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC   CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC   or send an email to license@isb-sdb.ch).
CC   CC
CC   EMBL: Z54141; CAA90835.1; -.
DR   SGD: S0004936; YMR317W.
KW   Hypothetical protein; Repeat.
SQ   SEQUENCE 1140 AA; 113070 MW; 0153BEC42FE5427 CRC64;

```

alignment_scores:

```

Quality: 186.00      Length: 1278
Ratio: 0.295         Gaps: 46
Percent Similarity: 49.374 Percent Identity: 18.310

```

alignment_block:

US-09-303-518D-651 x YM96_YEAST ..

Align seg 1/1 to: YM96_YEAST from: 1 to: 1140

5/5 AGCGTGTCCGATCGGCTCAGACACCACTATTGGCTTATGATGATGAC 624

```

|||||.....
24 SerThrThrSerThrThrThrProThrThrThrSerThrThrSerThrTh 40
625 AAACGCGCGATTATCTACTCGGCCATGGTTAATGGCGGCATAC 674
|||||.....
40 rSerThrLysValThrThrSerProGluIle.....IleValSerSers 55
675 ACATATGACGGGTGGGGAATTAATGGCTTANTTACTTTGAGCGGATG 724
|||||.....
55 eSerThrLeuValSerSerValValProGluPheThrSerSerSers 71
725 TCGGCCATGCCAGACTATGCGCCTATGCCGATTCGACGGTGGCAGGC 774
|||||.....
72 LeuSerSerAspThrIleAlaSerIle.....LeuSerSerGluSerLe 86
775 GACAGCGGTGCGCAATGTTATTATATACAACAACAATAAATAGCT 824
|||||.....
86 uValSerIlePheSerSerLeuSerTyrThrSerSerPheSerSert 103
825 GCTCAACGAGTTTACAAACGCGCTACCTTATTCGCGCAGG..... 868
103 hrSerValAsnAspValGluSerSerThrSerGlyProSerAsnSerTyr 119
869 AAACGCGTTTCCAGCTGATACGCAAAAGATTGGTTCTAGATGACATTAC 918
|||||.....
120 SerAlaLeuSerSerThrAsnAlaGlnLeuSerSerSerThrThrGluTh 136
919 AGAGCGGATACACATACGCTGTTTGTGAAACGCGCAGTAAACGACATT 968
|||||.....
136 rAspSerIleSer.....SerSerAlaIleGlnThrSers 148
969 TTCCTTT..ACATTCACAACAACAGGTAGCGGTACGCTAACAGAACAC 1017
|||||.....
148 erProGlnThrSerSerSerSerSerSerSerSerSerSerSerSerSer 164
1018 GAAAGGTTTCCAAATCCAAAGCTTAAAGTACAGACAGTCCGCTGTTGA 1067
|||||.....
164 uGlyLysSerSerValLeuGlnThrThrAlaSerSerSerAsp..... 178
1068 CGAATCTTTGAATGAACATGATTAAGAACCACTTACGGCGGAGGGGTG 1117
|||||.....
179 .....ThrThrAlaVal 182
1118 TTAATCAGTACCGTCCACAGTTAAACACGCTGAACACCTT...CTTT 1164
|||||.....
183 ThrSerSerThrPheThrThrLeuThrAspValSerSerSerProLysI 199
1165 ATGCATTTACGGCAACGCGCAACGTCATCTATATCAACACATCAACCAAG 1214
|||||.....
199 eSerSerSerGlySerAlaValThrSerValGlyThrThrSerAspAlas 216
1215 CCGCGCGCGTTCGTATTTGAAGGTGATTTTACGGTTCGCTGAAACA 1264
|||||.....
216 eTyrLysGluValPheSerSerSerThrSerAspValSerSerLeuSer 232
1265 ACCGAACGTGGCAAGCGCGCGCTTCATATCAATGAACACAGTACCGTT 1314
|||||.....
233 SerThrSerSerProAlaSerSerThrIleSer...GluThrLeuProH 248
1315 ACTTGGAAAGTAAACGCGTGGCAACGACCGCTGTCCAAAATGCGGCAA 1364
|||||.....
248 eSerSer.....ThrIleLeuSerIleThrSerSerProValSerSerc 263
1365 AGCAGCGTGCACGTTCAAGCCAAAGGGAACCAAGGCTGCATCAGCG 1414
|||||.....
263 LuAla.....ProSerAlaThrSerSerSerValSerSer 274
1415 TGGCGCAGCGTACAGTCATTTTGATCAGACGACGACAGCATAAAGCAA 1464
|||||.....
275 GluAlaSer.....SerSerThrSerSerSerValSerSe 286
1465 AAACAAGCCTTTAGTGAATCGGCTGNTCAGCGCGCAGGAGGTACGGTGA 1514
|||||.....

```



```

416 ugiylglnhslileglyleuthrpsp..... 424
416 AACAGGCTTTAGTGAATCGCTTGTTCAGCGGAGGGGTACGGTCAAA 1515
425 .....PheValIsglyglYAlaLeuAlaIaIaInglyThrleuthr 438
1516 CTG.....AATGCCGATAATCAGTTC.....AACCCGACAACTCTA 1553
439 LeuArgAsnAsnAlaValAlaGlnCysValLysAsnThrSerLysThrH 455
1554 TTTCGGC...TTTCGGCGGCGAGCTTTGGATTAAAC..... 1587
455 sgiylgYAlaIleleuAlaIleglyThrValAspLeuAsnGluThrIleSerg 472
1588 .....GGGCATTGCTTTGCG 1602
472 LuValAlaPheLysGlnAsnThrAlaLeuThrGlyAlaLeuSer 488
1603 TTCCACCGT.....ATTCAAAATACCGATGAAGGGCGGATGTCNCA 1646
489 AlaAsnAspLysValIleIleAlaAsnAsnPhcGlyLulleuPheG 505
1647 TCATAATGCCAACAACACATCCACCGTTACATT.....ACAG 1684
505 uGlnAsnGluValArgAsnHISglYAlaIleIleTyrcysglYcysArgS 522
1685 GGAATGAAGTATTACACA...CCGAGTGTGAAGATATCATACACTT 1731
522 eRAsnProLysLeuGlnLysAspSerGlyLysAsnIleAsnIleIle 538
1732 AATACAGCAAGAATTCGCTCAACAGCGTTGGTTGGCGAAGAATAC 1781
539 GlyAsnSerGlyAlaIle.....Th 545
1782 GACCAAAACAGACGGCGGCTCAACCTTTGTTACACCGCGCGAGAAG 1831
545 rPheLeuLysAsnLysAlaSerValLeuGluValaMetThrGlnAlaGlu 562
1832 ACCGCGCCGCTGCTTCGCGGAGCAAAATTTAAAGGGCAAC..... 1875
562 sp.....TyrAlaGlyglYAlaLeuTrpGlyHisAsnVal 574
1876 ..ATCACGCAACAACAGCGCAAACTGTTTTCAGCGGAGCGAGCACCC 1922
575 LeuLeuAspSerAsnSerGlyAsnIleGlnPheIleGly..... 587
1923 GCACGCTACATCATTTAGAGAACGGGTGCTCAAAATGGAAGTATCC 1972
588 .....AsnIleGly..... 590
1973 CACAAGGAATCGTGTGGACAACAGATGATCNACCGACGTTTAAA 2022
591 .....GlySerThrPheTrpIleGlyglYAlaIle..... 600
2023 GCGGAAATTTTCATATTCAGGGCGGCGGCGGTGATTCCCGCAATGT 2072
601 .....GlyglYAlaIleLeuSerThrAspAr 610
2073 TGCCAAAGTGAAGCGCATTCGNCATTTGAGCAATCAGCCCAACACTTT 2122
610 gValThrIleSer.....AsnAsnSerGlyAspValAlp 622
2123 TT..... 2124
622 helysglYAsnLysglYcInCysLeuAlaGlnLysTyrrValAlaProGln 638
2125 GGTCGCGACCGCATCAAGCCATACATCTGTACACGTTGCGAGCTGAC 2174
639 GluThrAlaProValGluSerAspAlaSerSerThrAsnLysAspGlu 655
2175 NGTCTGCAAAATGT..... 2190

```

```

655 sSerLeuAsnAlaCysSerHisGlyAspHisIleTyrrProProLysThrValG 672
2191 .....GTCAANAANAACATTTACCGACGATTAAGATGATGCTTCATG 2232
672 LuGluGluValProProSerLeuLeuGlnHISProValValSer 688
2233 ACTAAGACGACNTNAGCGGCAATGTCNACTNACNTNACNTNACTNNTN 2282
689 ThrAspIleArgGlyglYAlaIleLeuAlaIleGlnHisIlePheIleTh 705
2283 AAANCTCNCNGCGCTGCGCNCNCACTNANGCAATCTTAGTCAAAATGCG 2332
705 rAspAsnThrGlyAsnLeuArgPheSerGlyAsnLeu...GlyglYcys 721
2333 ATACAGCTTATACGTC..... 2349
721 LuGluSerSerThrValGlyAspLeuAlaIleValGlyglYAlaLeu 737
2350 .....AG 2351
738 LeuSerThrAsnGluValAsnValCysSerAsnGlnAsnValAlaPheSe 754
2352 CCACAACCGCCACCCCAAAACGCG..... 2373
754 rAspAsnValThrSerAsnGlyCysAspSerGlyglYAlaIleLeuAla 771
2374 .....AACCTTAGCGCTGCGGCAATGCG 2397
771 YsLysValAspIleSerAlaAsnHisSerValGluPheValSerAsnGly 787
2398 CAAGCAACATTTAATCAAGC.....ACATTAACGGCAACNCATCGGN 2441
788 SerGlyLysPheGlyglYAlaValCysAlaLeuAsnGluSerValAsnI 804
2442 TTTCGGCAATGCTCA...TTTAATCTAAGCAACAC..... 2475
804 eThrAspAsnGlySerAlaValSerPheSerLysAsnArgThrArgLeuG 821
2476 .....GCCGCAAAACAGCGAGCTGTGACGCTTTCGACCAAC 2511
821 YglYAlaGlyAlaIleAlaProGlnGlySerValThrIleCysGlysn 837
2512 GCTAAGCAACGTAAGCATTCGCGACTCAACGGCATGTCTCCCTACG 2561
838 .....GlnGlyAsnIleAlaPheLys 844
2562 CGATAAGCATATTCATTTGAAAACAGCGCG..... 2595
844 sGluAsnPheValPheGlySerGluAsnGlnArgSerGlyglYAlaI 861
2595 ..... 2595
861 leIleAlaAsnSerSerValaIleGlnAspAsnAlaGlyAspIleLeu 877
2596 .....TTTACCGGACA 2606
878 PheValSerAsnSerThrGlySerTyrrGlyAlaIlePheValGlySe 894
2607 ACTC.....ACGGCA 2617
894 rLeuValAlaSerGluGlySerAsnProArgThrLeuThrIleThrGly 911
2618 GCAAGGAAACGACATTACCTTAAAGACAGCGAATGACG..... 2658
911 snSerGlyAspIleLeuPheAlaLysAsnSerThrGlnThrAlaIaSer 927
2658 ..... 2658
928 LeuSerGluLysAspSerPheGlyglYAlaIleTyrrThrGlnAsnLe 944
2659 .....CTGC 2662
944 uLysIleValLysAsnAlaGlyAsnValSerPheTyrrGlyAsnArgAlaP 961

```



```

1491 LUGLNGLY.....LYSGLVAlaArg 1497
4096 TCATTCAACCGCGGCACATNTCCATCNCCTTATTNAGCTTC 4145
1498 ThrPheGlnYsThrArgPheGlnValAlaIleProPheGlyPheAl 1514
4146 CTATACCGATGCCGCTCGGGCAAGTCCGAACAGCGCTCAATACCGGNG 4195
1514 LeuGlnHisAlaIleYsSerArgGlySerArgAlaGlnValAsnSerValG 1531
4196 TATTGGCTCAGGATTCGCGAAACCCGACGCG..... 4230
1531 InLeuAlaIleValPheAspValYsArgGlyProValSerLeuIle 1547
4231 .....GAATGGCGCGTAACCGCGCA 4250
1548 ThrLeuYsAspAlaIleYsSerTrpYsSerYsGlyValAlaIlePr 1564
4251 AATCAAGGTTTCACGCTGTCNCTCCACGCTGCCGCCCAAGNCCGC 4300
1564 OCysLysAlaTrp.....LysAlaIleYsSerAsnAsnThrG 1577
4301 AACTGGAAGCGCAACAGCGCGGCATCAATTAGCTACCGCTGG 4347
1577 IuTrpAsnSerYsLeuSerThrYsLeuAlaIlePheAsnTrpGluTrp 1592

seq_name: SwissProt_40:120K_RICRI
seq_documentation_block:
ID 120K_RICRI STANDARD; PRT; 1300 AA.
AC P14914;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 120 kDa surface-exposed protein.
GN P120.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiinae; Rickettsia.
OX NCBI_Taxid=783;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=R;
RX MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT "Cloning, expression and sequence analysis of the gene encoding the
RT 120 kD surface-exposed protein of Rickettsia rickettsii.";
RL Mol. Microbiol. 3:1579-1586(1989).
CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS
CC RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -1- MISCELLANEOUS: ITS C-TERMINUS POTENTIALLY MAY BEAR THE EPITOPES
CC -1- CONFERRING ANTIGENICITY TO THE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC .....
DR EMBL: X16353; CAA34402.1; -
DR PIR: S07575; S07575.
DR InterPro: IPR003858; romPA_ompB.
DR Pfam: PF02708; romPA_ompB; 1.
KM Antigen: Glycoprotein; Cell wall; S-layer.
FT CARBOHYD 66 66 N-LINKED (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (POTENTIAL).

```

```

FT CARBOHYD 103 103 N-LINKED (POTENTIAL).
FT CARBOHYD 147 147 N-LINKED (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (POTENTIAL).
FT CARBOHYD 424 424 N-LINKED (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (POTENTIAL).
FT CARBOHYD 436 436 N-LINKED (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (POTENTIAL).
FT CARBOHYD 515 515 N-LINKED (POTENTIAL).
FT CARBOHYD 547 547 N-LINKED (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (POTENTIAL).
FT CARBOHYD 655 655 N-LINKED (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (POTENTIAL).
FT CARBOHYD 799 799 N-LINKED (POTENTIAL).
FT CARBOHYD 800 800 N-LINKED (POTENTIAL).
FT CARBOHYD 826 826 N-LINKED (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (POTENTIAL).
FT CARBOHYD 861 861 N-LINKED (POTENTIAL).
FT CARBOHYD 879 879 N-LINKED (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (POTENTIAL).
FT CARBOHYD 1116 1116 N-LINKED (POTENTIAL).
FT CARBOHYD 1128 1128 N-LINKED (POTENTIAL).
FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
FT CARBOHYD 1146 1146 N-LINKED (POTENTIAL).
FT CARBOHYD 1211 1211 N-LINKED (POTENTIAL).
SQ SEQUENCE 1300 AA; 132801 MW; E09B52C3F647243D CRC64;

```

alignment_scores: Quality: 185.00 Length: 1334
Ratio: 0.322 Gaps: 58
Percent Similarity: 43.103 Percent Identity: 18.441

alignment_block:

us-09-303-518d-651 x 120K_RICRI ..

Align seg 1/1 to: 120K_RICRI from: 1 to: 1300

```

1138 TTAACAACGGTGAACCTTTCTTTATCGATTACGCAAC..... 1179
      :::::  :::::  :::::  :::::  :::::  :::::
37 IleThrGlnAspSerAsnPheGlyAsnThrAspPheGlyAsnLeuAlaI 53
1180 .....GCGAACTCATCTTATCAACAACATC..... 1206
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
53 aglnlleLysValProAsnAlaIleThrLeuThrGlyAsnPheThrGlyA 70
1207 .....AACCAAGC.....GCGGCGGTTTGATTTGAAGGTGAT 1242
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
70 spAlaserAsnProGlyAsnThrAlaGlyValIleThrPheAspAlaAsn 86
1243 TTTACGGTCTCGCCTGAACCAAGACGAGGCGGCGCTTCA 1292
      ||:::  ||:::  ||:::  ||:::  ||:::  ||:::
87 GlyThrLeuGlnSerAlaSerAlaAsp.....AlaAsnValAl 99
1293 TATCACTGAAGACAGTACCTTACTTGGAAGTAACGCGTGGCAACG 1342
      :::::  :::::  :::::  :::::  :::::  :::::
99 aValThrAsnAsnIleThrAlaIle...GluIaSerGlyAlaGlyVal 115
1343 ACGCGCTGCCAAATCGCGCAAGACGACGCTGACGTTCAACCCAAAGG 1392
      :::::  :::::  :::::  :::::  :::::  :::::
115 aLeuLeuSer.....GlyThrHisAlaIleGluLeuThrGlu 127
1393 GAATCAACAGCTCGATC...AGCGTGGCGACGATACGATCTTTGGA 1439
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
128 GlyAsnAlaGlySerIlePheLeuAlaAspGlyThrValIle..... 142
1440 TCAGCAGCGACAGATAAAGCAAAAACAGCCCTTACTGAAATCGGCT 1489
      :::::  :::::  :::::  :::::  :::::  :::::
143 .....AsnGlyValAsnGlnThrAlaLeuValGlyG 154

```



```
1490 TGNTCACGGCGAGGGTACGTCGTCACATGCAATCCGATTAATCATGTTCAAC 1539
      ::::: |||||::: |||||:::
154 lYAlaLeuAlaAlaGlyThrIleThrLeuAspGlySerAlaThrIleThr 170
1540 CCCGAC..... 1545
      |||
171 GlyAspIleGlyAsnAlaGlyAlaAlaAlaLeuGlnArgIleThrIle 187
1546 .....AACCTCTATTTCG 1558
187 uAlaAsnAspAlaLysLysThrLeuThrLeuGlyAlaAlaAsnIleIleG 204
1559 GCTTTCGGCGCGGACGCTTGGATTTA.....AACGGCATTCGCTTTCG 1602
      || ||||| |||||::: |||||:::
204 lYAlaGlylYglYThrIleAspLeuGlnAlaAsnGlylYThrIleLys 220
1603 TTCACACCGATTCAAAAT..... 1620
      ::::: |||||
221 LeuThrSerThrGlnAsnAsnIleValAlaAspPheAspLeuAlaIleAl 237
1621 .ACCGATGAAGGGCGATGATGNCNATCATATGTCACAAACACATGCCA 1669
      |||||::: ::::: |||||:::
237 aThrAspGlnThrGlyValAlaAspAlaSerSerLeuThrAsnAlaGlnT 254
1670 CCGTTACCATTAACAGGAAGTAAGTATTACACAACCGAGTGTGAAGAT 1719
      ||::: |||||::: |||||:::
254 hrLeuThrIleAsnGlyLysIleGlyThrIleGlyAlaAsnAsnLysThr 270
1720 ATCATTAACACTTAATTAC.....ACGAACAATAATGGCTTCAACGGTTG 1763
      ||::: |||||::: |||||:::
271 LeuGlyGlnPheAsnIleGlySerSerLysThrValLeuSerAsnGlyAs 287
1764 GTTTCGCGAGAAAGAT..... 1779
      ::::: |||||
287 nValAlaIleAsnGlnLeuValIleGlyAsnAspGlyAlaValGlnPheA 304
1780 .....ACGACCAAAACGACGAGCGGCGTC 1803
      |||||::: |||||:::
304 lAhAspThrTyrlLeuIleThrIleThrIleThrAsnAlaAlaGlyGlnGly 320
1804 AACCTGTTTACAGCCGCGCGAGAAAGACCGCACCCGCGCTTTCGCG 1833
      ||::: |||||::: |||||:::
321 LysIleIlePheAsnProValAlaAsnAsnGlyThrThr...LeuAlaAl 336
1854 CGAACAATAATTAAACGC..... 1872
      ::||| |||||:::
336 aGlyThrAsnLeuGlySerAlaThrAsnProLeuAlaGlnIleAsnPheG 353
1872 ..... 1872
353 LysSerLysGlyValAsnValAspThrValLeuAsnValGlyGlnGlyVal 369
1873 .....AACATCAGCAAAACAAC.....GCCAAGCT 1898
      |||||::: |||||:::
370 AsnLeuTyrlaThrAsnIleThrThrThrAsnValGlySerPhe 386
1899 GTTTTACAGCGGACCGACGACGACGACGCTACATCATTTAGAGAGCG 1948
      ||::: |||||::: |||||:::
386 eValPheAsnAlaGlylYThrAsnIleValSerGlyThrValGlylYlY 402
1949 GGTGTCAAAAATGGAAGGTATCCACACAGAGAATAATCGTGTGGACAC 1998
      ::::: |||||::: |||||:::
403 .....GlnGlnGlyAsnLysPheAsnThrValAlaIleLeuGlnAsn 415
1999 GACTGATCNACCGC.....ACGTTTAA..... 2022
      ::::: |||||:::
416 GlyThrThrValLysPheLeuGlyAsnAlaIleThrPheAsnGlnLysnThr 432
2023 ....GCGAAAATTTCCATATTCAGGGCGGACGCGGTATTTCCGCA 2068
      ||| ||| |||||::: |||||:::
432 rIleAlaAlaAsnSerThrLeuGlnIleGlylYAsnTyrlThrAlaAspC 449

2069 ATGTGCCAAAGTGAAGCGAT..... 2091
      |||||::: |||||:::
449 ySValAlaSerAlaAspGlyThrGlylYValGlnIlePheValAsnThrGly 465
2091 ..... 2091
466 ProIleThrValThrLeuAsnLysGlnAlaAlaProValAsnAlaLeu 482
2092 .....TGNC 2095
482 sGlnIleThrValSerGlyProGlyAsnValAlaIleAsnGlnIleGlyA 499
2096 ATTGACCAATCAAGCCAGAGTTTGTGTGGCGACCGCATTCAAAGC 2145
      ::::: ||||| |||||:::
499 snAlaGlyAsnHisHisGlyAlaValThrAspThrIleAlaPheGlnAsn 515
2146 CATCAATC..... 2154
      ::::: |||||
516 SerSerLeuGlyAlaValAlaPheLeuProArgGlylYleProPheAsnAs 532
2155 .....TGTAACAGCTTCGACGTGACNGCTGTGACAA 2185
      |||||::: |||||:::
532 pAlaGlyAsnThrMetProLeuThrIleLysSerThrValGlyAsnLysT 549
2186 ATTGT.....GTGCAANA 2199
      ::::: |||||:::
549 hrAlaLysGlylYPheAspValProSerValAlaValLeuGlyValAspSer 565
2200 ANCATTAACGAGATAAAGTATGCTTCATTGACTAAGACNGACNTNAG 2249
      |||||::: |||||:::
566 ValIleAlaAspGlyGlnValIleGly.....AspGlnAs 577
2250 CGGCANTGTAGACTNNCCNATNACGNTNNTTAAANCTC..... 2289
      |||||::: |||||:::
577 nasnIleValGlyLeuGlyLeuGlySerAspAsnGlylYleIleValAsnA 594
2290 .....NCGGGCMT 2298
594 lathrThrLeuTyrlAlaGlylYleSerThrLeuAsnAsnGlnGlylYnT 610
2299 GCNNCACTNNAANGCAATCTTACTGCAAAATGGGATACAGTTAT..... 2343
      |||||::: |||||:::
611 ValThrLeuSerGlyGlyValProAsnThrProGlylYThrValTyrlYle 627
2343 ..... 2343
627 uGlylYnThrGlylYleGlyAlaSerLysPheLysGlnValThrPheThrThA 644
2344 .....ACAGTCAGCCACACGACGACCCCAAAACGCGC 2373
      ::||| |||||::: |||||:::
644 sPTyrAsnAsnLeuGlyAsnIleIleAlaThrAsnAlaThrIleAsnAsp 660
2374 AACCTTAGCCTGCTG.....GGCAATGCCCAAGCAACATTTAATCAAGC 2417
      ::||| |||||::: |||||:::
661 GlyValThrValThrThrGlylYlYleAlaGlylYleGlylYlYlYlYlY 677
2418 CACATTAACGCAACNCAFCGNTTCGGGCAATGCTTATTAT..... 2463
      |||||::: |||||:::
677 sIleThrLeuGlySerValAsnGlyAsnGlylYlYlYlYlYlYlYlYlY 694
2464 ....CTAAGCAACAACGCGGACCAAAACGAGCTGACGCTTTCGAC 2508
      |||||::: |||||::: |||||:::
694 lYlYleuSerAsnSerThrSerMetIleGlylYThr...ThryslAlaAsn 709
2509 AACGCTAAGCAACGTAAGCCATTCGACATCAACGCAATGCTCCCT 2558
      |||||::: |||||::: |||||:::
710 AsnGlylYThrValThrTyrlYlYlYlYlYlYlYlYlYlYlYlYlYlYlY 726
2559 AGCCGATTAAGCAGCATTCATTGAAACAGCGCGCTTACCGGACAC 2608
      ::||| ||| |||||::: |||||:::
726 pSerAspThrProVal.....AlaSerValArgPheThrGlySerA 740
2609 TCAGCGGCGACAGGAAACAGCA..... 2631
```



```

seq_name: swissprot_40:WAPA_BACSU
seq_documentation_block:
ID WAPA_BACSU STANDARD: PRT: 2334 AA.
AC 007833:
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Wall-associated protein precursor.
GN WAPA OR N17G.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=93302506; PubMed=8316082;
RA Foster S.J.;
RT "Molecular analysis of three major wall-associated proteins of
RT Bacillus subtilis 168: evidence for processing of the product of a
RT gene encoding a 258 kDa precursor two-domain ligand-binding
RT protein."
RL Mol. Microbiol. 8:299-310(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=95219088; PubMed=7704263;
RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
RT genome containing the hut and wapa loci."
RL Microbiology 141:337-343(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / BGSC1A1;
RX MEDLINE=97124196; PubMed=8969509;
RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
RA Mwa Y., Fujita Y.;
RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
RT containing the lig and cel loci, and creation of a 177 kb contig
RT covering the gut-sacx region."
RL Microbiology 142:3113-3123(1996).
CC -1- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
CC MOTILITY, SECRETION OR DIFFERENTIATION.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
CC INTO THE MEDIUM.
CC -1- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS: THE N-TERMINUS. HAS THREE
CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
CC MOTIF REPEATED 31 TIMES.
CC -1- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
CC SIMILARITY TO THE REPEAT IN E. COLI RBS GROUP OF PROTEINS (RBSA-D).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L05634; AAA22883.1; -
DR EMBL; D31856; BAA06656.1; -
DR EMBL; D29985; BAA06260.1; -
DR EMBL; D83026; BAA11683.1; -
DR EMBL; Z89124; CAB15959.1; -
DR PIR; S32920; S32920.
DR Subtilisin; BG10797; WAPA.
DR InterPro: IPR003305; CBD_6.
DR Pfam: PR02018; CBD_6; 1.
KW Cell wall; Repeat; Signal; Complete proteome.
FT SIGNAL 1 28 OR 32 (POTENTIAL).

```

```

FT CHAIN 29 2334 WALL-ASSOCIATED PROTEIN.
FT DOMAIN 504 869 3 X 101 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 504 605 1-1.
FT REPEAT 606 736 1-2.
FT REPEAT 737 869 1-3.
FT DOMAIN 1021 2139 31 X 21 AA APPROXIMATE TANDEM REPEATS OF
FT REPEAT 1021 1040 X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).
FT REPEAT 1041 1061 2-1.
FT REPEAT 1062 1082 2-2.
FT REPEAT 1083 1102 2-3.
FT REPEAT 1103 1128 2-4.
FT REPEAT 1129 1148 2-5.
FT REPEAT 1149 1168 2-6.
FT REPEAT 1169 1188 2-7.
FT REPEAT 1189 1207 2-8.
FT REPEAT 1208 1227 2-9.
FT REPEAT 1228 1247 2-10.
FT REPEAT 1248 1267 2-11.
FT REPEAT 1268 1287 2-12.
FT REPEAT 1288 1307 2-13.
FT REPEAT 1308 1327 2-14.
FT REPEAT 1328 1347 2-15.
FT REPEAT 1348 1367 2-16.
FT REPEAT 1368 1387 2-17.
FT REPEAT 1388 1407 2-18.
FT REPEAT 1408 1427 2-19.
FT REPEAT 1428 1447 2-20.
FT REPEAT 1448 1467 2-21.
FT REPEAT 1468 1487 2-22.
FT REPEAT 1488 1507 2-23.
FT REPEAT 1508 1527 2-24 (APPROXIMATE).
FT REPEAT 1528 1547 2-25.
FT REPEAT 1548 1567 2-26.
FT REPEAT 1568 1587 2-27.
FT REPEAT 1588 1607 2-28.
FT REPEAT 1608 1627 2-29.
FT REPEAT 1628 1647 2-30.
FT REPEAT 1648 1667 2-31.
SO SEQUENCE 2334 AA; 258329 MW; B75138CCD278BA3 CRC64;

```

```

alignment_scores:
Quality: 184.00 Length: 1242
Ratio: 0.333 Gaps: 65
Percent Similarity: 44.444 Percent Identity: 18.680

```

alignment_block:

US-09-303-518d-651 x WAPA_BACSU

Align seg 1/1 to: WAPA_BACSU from: 1 to: 2334

```

160 TATCGGCACTTGGCGAAATAAGCAAGTTGCGAGTGGCGGCAAGA 209
||| :||||| :|
988 TyleuGlndeuThrAlaSpGlnPheLeuLysThrLysAs 1004
210 T..... 210
1004 pGlnThrAsnAlaTyPheAsnLysLysGlyLysLeuGlnValy 1021
211 .....ATTGAGCTWTACAACAAAAAGGGGAG 237
1021 aAspGlyHisAsnAlaThrValTyThrTyAsnAspLysAsnGln 1037
238 TTGGTGGCAATCAATGACAAAGCCGATGATTGTCGTGCT 287
||||| :| :|
1038 LeuThrAlaIleThraSpAlaSerGlyAlaGlyLeuThrPheThrTyAs 1054
288 GTGCGGTAAGCGCGTGGCGGCAATGCTGGGCGC..... 318
||||| :| :|
1054 pGluAsnGlyHisValThrSerIleThrcLysProlLysAsnLysValT 1071
319 .....GATCAATATATTGAGGCGGACAT...AACGGC 351
||||| :| :|

```

```
1071 hTyrSerTyrGluAsnSpleLeuLysValThrAspThrAspGly 1087
352 GCGTATACAAACGTTGATTTGGTCGGAAGAAAGNAATCCGATACACA 401
1088 ThrValThrSerTyrAspTyrAspSerGly 1098
402 CCGTTTCTTACCAATGTGAAAAAATAATATATAAGCCTGACANT 451
1099 .....ArgLeuValLysSerIleAsnSerThrGluA 1111
452 CACACCCCT.....TACACGGCGGATTATCATGCGCGTTTG 489
1111 IalysProValPheThrGluTyrGlnTyrSerGlyHis.....ArgLeu 1125
490 CATAAATTTGTCACA.....GATGC 509
1126 GluLysAlaIleAsnAlaLysLysGluThrTyrValTyrSerTyrAspAl 1142
510 AGAA..... 513
1142 aaSplyLysThrLeuLeuMetThrGlnProAsnGlyArgLysValGlnT 1159
514 .....CCTGTGAAATGACGAGTACATGATGAG 540
1159 YrGlyTyrAsnGluAlaGlyAsnProIleGlnValIleAspAspAlaGlu 1175
541 GGAATACCTATTCCGATAAAGAAATAT..... 570
1176 GlyLeuLysIleThrThrAsnThrLysTyrGluGlyAsnAsnValAla 1192
571 .....CCGAGCGGTCCGATCGGCTACGACACCATATATGGC 610
1192 uasPvalAspProAsnAspValGlyThrGlyLysAlaThrGlnSerTyrG 1209
611 GTTATGATGATGACAAACACGCGGATTTA...TCTACTCCGGCGCATG 657
1209 IntYrAspLysAsp.....GlyAsnValThrSerValLysAspAlaTyr 1223
658 TTATTTGGCGCAATACACATATGACAGCGTTGGAAATATAGCGGTANT 707
1224 .....GlyThrGlnThrTyrGluTyrAsn...LysAsnAsnAspValTh 1237
708 TAGTTGAGCGGCGATGTGCCCATGCGCAACGAC.....TATGGCCCA 751
1237 rLysMetLysAspThrGluGlyAsnValThrAspIleAlaTyrAspGlyL 1254
752 TGCCGATTCAGAGTGGCGGACGACGCGT.....TGCCCAATGTTT 795
1254 euAspAlaValSerGluThrAspIleSerGlyLysSerSerSerAlaAla 1270
796 ATTATGACAAACAAACAAATTAATGCTGCTCAACGAGTTTACAAAC 845
1271 ValTyrAspLysTyrGlyAsnGln.....IleGlnSe 1281
846 CGGCTACCTTATTCGCGACGGAAACGGTTTCCAGCTGATACGCAAG 895
1281 rSerLysAspLeuSerAlaSerThrAsn.....IleLeuLysA 1294
896 ATTGCTTACGATGACATTTACAGAGC..... 924
1294 spGlySerPheGluAlaGlnLysSerGlyTyrPasnLeuThrAlaSerLys 1310
925 GATACACATACCGTCTNTTTTGAACCCGACGATACGACATTTTTC.. 972
1311 AspArgArgLysIleSerValIleAlaAspLysSerGlyValLeuSerG 1327
973 .....TTTACATCCACACACA 988
1327 ySerLysAlaLeuGluValLeuSerGlnSerThrSerAlaGlyThrAsp 1344
989 ACGGTACCGGTACGTAACAGAAACAAAGGTTNCCATGCAATCCA... 1035
1344 isGlyTyrSerSerAlaThrGlnThrValGlnLeuGluProAsnThrThr 1360

1036 .....AAGCTTAAAGTACAGACAGTCCA.....CT 1061
1361 TyrThrLeuSerGlyLysIleLysThrAspLeuAlaLysSerArgAlaTyr 1377
1062 GTTTGACGAATCTTTGAATGAACCTGATAAAGAACAGCTTTAC..... 1104
1377 rPheAsnIleAspLeuArgAspLysAspGlnLysArgIleGlnTrrPhe 1394
1105 .....GCGGACGGGCTTAAATGATACGCTCCAGGTTA 1140
1394 IsaGlnLysTyrSerAlaLeuAlaGlyLysAsnSpirThrLysArg... 1409
1141 AACACGGTGAACCTTTCTTATGATATACGCAAC...GGCAACT 1187
1410 .....GlnIleThrPheThrThrProAlaAsnAlaGlyLysAl 1422
1188 CATCTTATCAAAACATCAAC.....CANGCGCGGCGGCTT 1225
1422 aValValTyrMetGluValAspHisLysAspLysAspGlyLysGlyLysA 1439
1226 TGATTTTGA.....GGTATTTTACGGTCTGC... 1254
1439 IatrrPheAspGluValGlnLeuGluLysGlyGlnValSerSerSerTyr 1455
1255 ...CCTGAAACAAACGAACGTGGCAAGCGCGGCGCTTATATCAATGA 1301
1456 AsnProValGlnAsnSerSerPheThrSerAla..... 1466
1302 AGACATACCGTTACTTGGAAGTAAACGGCGGTGCA..... 1338
1467 .....ThrGluAsnThrPasnValSerGlyAlaSerValAspSerGlu 1481
1339 .....AAGACCGCTGTCCAAATCGCAAAAGCAGCGCTCACAGT 1380
1481 LuGlyPheAsnAspAspValSerLeuLysAlaIleAlaArgThrSerAlaSer 1497
1381 CAAGCCAAAGGGGAAACCAAGCCTGATCAAGCGTGGCGGCGGTACAGT 1430
1498 GlnAlaGly.....SerValThrLysGlnThrVa 1507
1431 CATTTTGATGACGCGGACAGATAA.....G 1459
1507 lValLeuGlnSerAlaAsnAspLysProValTyrIleThrLeuThrG 1524
1460 GCAAAAACAAAGCCTTAGTGAATCGGCTTGNTCAGCGGAGGGGTACG 1509
1524 LysMetSerLysAlaSerSerVal..... 1531
1510 GTGCAACTGAATGCCGATATCATGTTCAACCCGCAACAACTCTATTGCG 1559
1532 .....LysPheThrAspGlnLysAspTyr.... 1539
1560 CTTTGGCGGCGGACGTTTGATTTAAACGGCGATTCGTTTCGTCACC 1609
1540 .....SerLeuGlnAlaAsn..... 1544
1610 GTATTCAAAATACGATGAAGGGCGATGTTGNCNATCATATATCCACA 1659
1545 ...ValThrTyrAlaAspGlySerThrGlyLe.....TyrAsnAlaLys 1558
1660 ACAACATTCACCGTTACCATTTACAGGAATGAAGATTTACACACGCG 1709
1559 PheProSerGlyThrGlnGluThrPasnArgAlaValAlaValIleProLy 1575
1710 TGGTAGATATCATATAGACTTAAT.....TACACCAAAAGAA 1747
1575 sThrLysProIleAsnLysValAspIleSerIleLeuPheGlnLysSerA 1592
1748 TTGCTTACAAAGCGTGTGGCGAG..... 1773
1592 IatThrGlyThrValTrrPheAspAspIleArgLeuIleGlnLysSerLeu 1608
```

```

1774 .....AAAGTACGACCAAAACGACGCGGCTC..... 1803
      |||:|||||
1609 LeuThrLysSerThrTyrAspSerAsnGlyAsnTyrValThrLysGlu 1625
      |||:|||||
1804 ....AACCTGTTACCAAGCCCGCCGAGAA...GACCGACCCNGCTGC 1846
      |||:|||||
1625 uAspGluLeuGlyTyrValThrSerThrAspTyrAspGluThrGlyLysL 1642
      |||:|||||
1847 TTTCGGCGGACCAAAATTTAAACGC..... 1872
      |||:|||||
1642 ystThrSerGluThrAspAlaLysGlyLysThrThrTyrThrTyrAsp 1658
      |||:|||||
1873 .....AACATCCGCAAAACAAACGCGCAACTGTTTT 1904
      |||:|||||
1659 GluAlaAspGluLeuThrAsnMetThrLeuSerAsnGlyThrSerIleLe 1675
      |||:|||||
1905 CAGCGGAGACCGACCGGACCGCTCATCTTTAGGAAGCGGGGCT 1954
      |||:|||||
1675 u.....HisThrAspLysGluGlyAsnGluValS 1686
      |||:|||||
1955 CAANAATGGAAGTATCCACAAGAGAAATCGTGGACAAACGACTGG 2004
      |||:|||||
1686 erLysThr.....IleArgAlaGly..... 1692
      |||:|||||
2005 ATCCACCGCACGTTTAAAGCGGAAATTTCCATATTCAGCGCGGAGGC 2054
      |||:|||||
1693 AlaAspGluThrTyrLysPheGlu...TyrAspValMetGlyLysLeuVa 1708
      |||:|||||
2055 GGTGATTTCCCGCATGTGCCAAAGTGAAGGCGAGTGCATTTAGCA 2104
      |||:|||||
1708 LysThrThrAspProLeuGlyAsnValIleuAlaSerGluTyrAspAla 1725
      |||:|||||
2105 ATCAGCGCCCAAGCAGTTTGTGTGCGACCGCATCAACCATACATC 2154
      |||:|||||
1725 sn..... 1725
      |||:|||||
2155 TGACACGTTGCGACGAGNGCTGCACAATGTGTGCANAANAANCAT 2204
      |||:|||||
1726 .....SerAsnLeuThrLysThrIleSerPro..... 1734
      |||:|||||
2205 TACGACGATAAAGTGATGCTTCATTTGACTAAGACNGACNNTAAGCGCA 2254
      |||:|||||
1735 AsnGlyAsnGluValSerLeuSerTyrAspGlyThrAspArg..... 1748
      |||:|||||
2255 NTGTNACNTNCCNATNACGNTNNTNAANCTCNCNGCGCTGCNCA 2304
      |||:|||||
1748 ..... 1748
      |||:|||||
2305 CTNAANGCAATCTTAGTCAATGCGGATACGCTTATACGATCAGCA 2354
      |||:|||||
1749 ...ValLysSerLysSerTyrAsnGlyThrGluLysThrIlePheThrTy 1764
      |||:|||||
2355 CAAGCGCACCCAAACGGCAAC...CTTAGCCCTGCGGCAACGCCCAAG 2401
      |||:|||||
1764 rAsp.....LysAsnGlyAsnGluThrSerValValAsnLysGluGln 1779
      |||:|||||
2402 CAACATTTAATCAAGCCACATTAAGCGCAACNCATCGNTGCGCAAT 2451
      |||:|||||
1779 snThrThrLysLysArgThrPheAspAsnLysAsn..... 1790
      |||:|||||
2452 GCTTCATTATCTAAGCAACAGCGCGCAAAACGCGCATGTGACGCT 2501
      |||:|||||
1791 .....ArgLeuThrGluLeuThrAspArgGlySerGlnThr 1804
      |||:|||||
2502 T.....TCGACAAACGCTAAGCA..... 2520
      |||:|||||
1804 pThrTyrProSerAspSerAspLysLeuThrPheSerTrpIleHisG 1821
      |||:|||||
2521 .....AACGTAAAGCATTCGCACTCAACGCGCAAT 2550
      |||:|||||
1821 LysAspGluLysGlyThrAsnGlnPheThrTyrAsnLysLeuAspGlnMet 1837
      |||:|||||
2551 GTCTCCCTAGCCGATAG...GCAGTATTCATTTT.....GAAA 2588

```

```

1838 IleGluMetLysAspSerThrSerSerTyrSerPheAspTyrAspGluAs 1854
      |||:|||||
2589 CAGCGGC.....TTTACCGGACAACTCAGCGGACGACGAGANA 2626
      |||:|||||
1854 ngLysAsnValGluThrPheIleThrGlyAsnGlyGlyThrSerPheS 1871
      |||:|||||
2627 CAGCATTAACCTTAAGACACGCAATGACGCTCCGCTCAGCGACGANA 2676
      |||:|||||
1871 erTyr.....AspGluArgAsnLeuValSerSerLeuHis 1882
      |||:|||||
2677 TTAGCAATTTAAACCTTGACAAACCCACCATTAATTCATTCGCGCTA 2726
      |||:|||||
1883 IleGlyAsp.....LysAsnGlyLysAspIleLeuThrGluSerTyr 1896
      |||:|||||
2727 TCGCCACGATGCTGCAGCGCGCGCAACCGGACGCTGACACACCGCC 2776
      |||:|||||
1896 rGluTyrAspAlaAsnGly..... 1902
      |||:|||||
2777 GCCGCGCTGCGCGCTTCCTATTATCCGTTACACCGCAACTTCGGTA 2826
      |||:|||||
1903 .....AsnArgThrThrIleAsnSerSerAlaSerGlyLysVal 1915
      |||:|||||
2827 GAATCCCGTTTCACACGCTGACGCTAAACGCAATTTGAC..... 2868
      |||:|||||
1916 GlnTyrGluTyr.....GlyLysLeuAsnGlnLeuVa 1926
      |||:|||||
2869 .....NCTCAAGACAA...TTCCGCTTTATTCGACACT 2902
      |||:|||||
1926 LysGluThrHisGluAspGlyThrValIleGluTyrThrTyrAspGlyP 1943
      |||:|||||
2903 TCGGCTACGACGACGACAAATGAGCTGCGGGAAGTTCCGAAAGNCT 2952
      |||:|||||
1943 heGlyAsnArgLysThrValThrThrIleLysAspGlySerSerLysThr 1959
      |||:|||||
2953 TACACCTTGCGGCTCAACATACCGGCAACGACCCGTAAGCTTCATCA 3002
      |||:|||||
1960 ValAsnAlaSerPheAsn.....IleMetAsnG 1969
      |||:|||||
3003 ATTGACGCTAGTCGAAGGAAAGACAAACACCGCTGCCGAACCTTA 3052
      |||:|||||
1969 nLeuThrLysValAsn.....AspGluSerIleS 1979
      |||:|||||
3053 ATTTCACCTCGCAAAACGACACGTCGATGCCGCGCGGCGGTACCAA 3102
      |||:|||||
1979 erTyrAspLysAsnGlyAsnArgThrSerAspGlyLysPheThrTyrThr 1995
      |||:|||||
3103 CTCATCCGCAAGACGCGGAGTTCGCTGCATAATCCGTCAAAGACA 3152
      |||:|||||
1996 TrpAspAlaGluAspAsn...LeuThrAlaValThrLysLysGluAspL 2012
      |||:|||||
3153 AGAGCTTTCGCAACACTCGGACGAGCAAGCAACCAAAACAGCGCGAA 3202
      |||:|||||
2012 yspProPheAlaThrTyrLysTyrAspGluLys...GlyAsnArgIleGln 2027
      |||:|||||
3203 AAGACACGCGCAACGCTTGACG 3226
      |||:|||||
2028 LysThrValAsnGlyLysValThr 2035
      |||:|||||

```

seq_name: SwissProt_40:VG37_BPT2

seq_documentation_block:

ID VG37_BPT2 STANDARD; PRT; 1341 AA.

AC P07067;

DT 01-APR-1988 (Rel. 07, Created)

DT 01-APR-1988 (Rel. 07, last sequence update)

DT 16-OCT-2001 (Rel. 40, last annotation update)

DE Long tail fiber protein p37 (Protein Gp37) (Receptor recognizing

DE protein).

GN 37.

OS Bacteriophage T2.

OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;

OC T4-like phages.


```
392 ..... 392
398 AlaGlyAspGlyGln.....
1464 AAACAAAGCCTTAGTGAATCGGCTGTCAGGCG..... 1500
393 .....ThriHisIleGlyTyrAsnSerGlyGlyLysMetSerH 405
1501 .....AGGGTACGGTGCACACTGATGCGGATTAATCAAG..... 1533
405 IStYrPheArgGlyLysGlyGlnThrAsnIleAsnThrGlnLysGlyMet 421
1534 ...TTCAACCCCGAC.....AAACTCTATTTCGGCTTCGCGGCGGACG 1574
422 GluValAsnProGlyIleLeuLysLeuValThrAspSerAsnAsnValG 438
1575 TTGGATTAAACGGGCAATTCGCTTCGTCACCGTATTCAAATACCG 1624
438 nPheTyrAlaAsnGlyThrValSerSerIleGlnArgIleLys...Phe 454
1625 ATGAAGGGCGGATGATGNCNATCATATATGACCAACACATCCACGGT 1674
454 spasnGly.....Leu 457
1675 ACCATTACAGGGAATGAAAGTATTACACACGAGTGTAGAAATATCAA 1724
458 ValLeuThrGly.....AlaArgProAspGlyIleGlnLeuAs 470
1725 TAGACTTATTACAGCAAAAGAAATTCCTACACGGTTGGTGGCAGA 1774
470 P..... 470
1775 AAGATAGACCAAAACGAAGGGCGGCTCAACCTGTTTACCAGCGGCC 1824
471 .....AlaProThr 473
1825 GCAGAAAGC.....CGCACCCGCTGCTTCCGCGGAAACAAATTTAAA 1868
474 AlaAlaAspGlyThrIleLeuThrAlaGlyThrArgAlaG 490
1869 CGGCAACATCAGCAAAACGCAAACTGTTTTCAGCGGCGACCGCA 1918
490 YGlnAsn.....LysSerTyrValSerIleLysAla 501
1919 CACGCAAGCCTACATCATTTAGAAAGCGGCTGTCAAAATAGAGGT 1968
501 rPGLYAsnSerPheAsn.....AlaSerGlyAspArgAlaArgGluThr 515
1969 ATCCCAAGAGAAATCGTGTGGACACGACTGATCNACCGCACGTT 2018
516 Val..... 516
2019 TAAAGCGGAAATTCATATTCAGGCGGCGACGCG.....GTGATTT 2062
517 .....PheGluValGlyAspGlyGlnGlyPheHisPheTyrS 529
2063 CCGCGCAATGTGCAAGTGAAGCGATGNCATTTGAGCAATCACGCC 2112
529 eRgInaGlyAlaProAlaProAlaProLysSerThr..... 539
2113 CAAGCACTTTTGGTGTGCGACCGCATCAAGCCATCAATCTGTACAG 2162
540 .....ValGlyProIleGlnLeuArgValAsn..... 548
2163 TTGGACTGCGCNGCTGTGACAATTTGTGCAANAANAANCATTAACG 2212
549 .....GlyGlyLeu.....LeuThrAlaG 555
2213 ATAAAGTATTGCTTCATTGACTAGACAGCAGCAGCAGCAGTAAAGN 2262
555 LysIleValAlaSerGlySerIleThrThrGlnSer..... 567
2263 CTNNCCATNATNCGTNNTTNAAANCCTCNCNGCGCTCNCNCACTNAA 2312
NGG 2312
568 .....SerLeuAsnValAlaAsn 573
2313 CAATTGATGGCAATAGCGGATACAGTATATACAGTACCGCAACGCCA 2362
573 nGlyLeuSerValAsnGlyAlaLysPheGlyGlyThrAlaAsnAlaL 590
2363 CCCAAAGCGCAACCTTAGC.....CTCGTGGCAATGCCCAAGCA 2403
590 euArgIleThrPasnAlaGluTyrGlyValIlePheArgArgSerGln 606
2404 ACATTTAATCAACGCAATTAACGGCAACNCATCGGNTTCGGCAATGC 2453
607 AsnPheTyrIleIleProThrAsnGlnAsnGlnGlyLysSerGly 623
2454 TTCATTATCTAAGC.....AACAAAGCCCGCA... 2481
623 eHisSerSerLeuArgProValArgIleGlyLeuAsnAspGlyAlaVal 640
2482 .....CAAAAGCGGAGTCTG...ACG 2499
640 LysLeuGlyArgAspSerPheIleValAspGlnAsnAlaLeuThr 656
2500 CTTTCGACAAACGCTAAGCAACGTAAGCATTCCGCACTCAACGCCAA 2549
657 IleAsnSerAsnSerArgIleAsnAlaAsnPheArgMetGlnLeuGly 673
2550 TGCTCCCTAAGCCGAT..... 2565
673 nSerThrTyrIleAspAlaGluCysThrAspThrValArgProAlaG 690
2566 .....AAGCAGATTTCCAT 2580
690 IArgIleSerPheValSerGlnAsnAsnGlnAsnValArgAlaProP 706
2581 TTGAA..... 2586
707 MetAsnIleAsnArgThrAspThrSerThrTyrValProIleLeuG 723
2587 .....AACAGCCGCTTTACG...GGACACTCAGCGGCA 2617
723 nArgTyrValGlnGlyAsnSerCysTyrSerLeuGlyThrLeuIle 740
2618 GCAAGANACAGCATTAACCTTA..... 2640
740 hreIysAspPheArgIleHisTyrHisGlyGlyIleAspAsnLys 756
2641 .....AAAGACAGCGAATG 2654
757 GlyProGlnLysAlaAspLeuAlaTrpGlnPheArgArgAspGly 773
2655 GACGTCGCGCTACAGCAGCAATTA.....G 2680
773 eArgSerProAsnLysIleGlnIleAsnAlaValThrIleGlyThr 790
2681 GCAATTTA.....AACCTGACACAGCCGACCATTAACATC 2715
790 LysAsnIleHreGlyGlyThrGlyAsnPheAlaAsnLeuAsnThr 806
2716 AATTCGCGCTATCGCACGATGCTGAGCGGCGCAACCGCGCAGNGT 2765
807 Asn.....ArgLysThrThrValGlyGlyThrAlaGlySer... 818
2766 AGACAGCGCGCGCGCGCTTCGCGCGCTTCCTATTA..... 2802
819 .....SerValValGlyTyrTyrLysP 826
2803 .....TCCGTTACACGCGCAACCTGCGTGAATCCGTTTCAACAGCG 2847
826 hreAlaThrValThrIleProGlnSer.....ThrGly 836
2848 ACGGTAAACGCAAAATTAAGCAGTCAAGACATTCGCTTATGTGCGA 2897
837 ThrValThrPheLysIleSerGlyAlaGlyAlaGlyPheAsnPheLys 852
```

```

2898 ACTCTGCGCTACCGAAGCAGCAATTTGAGTGGCGGAAGTTCCGAG 2947
853 .....TyrasnGlnAlaSerIleAlaGlu..... 860
2948 GNACTTACACCTTGCGGTCAACATACCGCAACGAGACC..... 2988
861 .....IleValLeuArgThrGlyAsnAsnProLysGlyIle 872
2989 .....GTAGCTCGATCAATTCAGCGT 3011
873 AsnAlaValLeuTrpAsnArgSerAspLeuSerPheAsnGlnIleAlaThr 889
3012 AGTGAAGGAAAGACAGACAAA..... 3033
889 rMetAsnThrSerAspThrTyrAspValTyrValPheCysGluGlyT 906
3034 .....CCGCTGCGGAAACCTTAATTTC 3057
906 yTrhAsnAlaLeuIleValGluTyrSerCysSerGluAsnSerSerVal 922
3058 ACCCTGCAAAAGCAAC.....GTGCGATGGCGGCGC 3089
923 ThrValValGlyLeuAsnGlyGlyIleGlnProValValAspAla..... 937
3090 GTGGCGTTACCACTCATCCGCAAGACGCGAGTTCGCGCTGCATATC 3139
938 .....LeuProGluGly.....HisValVal 944
3140 CGGTCAAGAACAGAGCTTTCGCAACAATCGCGCAAGCAAGCCAAA 3189
944 aGlyIleThrValArgLeuLeuAsnAsnIleGlyGlyThrPheAlaAla 960
3190 AAACAGGCGGAA..... 3201
961 GlyIleSerAspIleValThrArgGlyGlyTyrValAlaAspAsnGlnIly 977
3202 .....AAAGACAACGCGCAAGCGCTTGACGC..... 3228
977 sGlyMetArgIleLysSerAsnGlyAsnAsnIleGlySerAsnAlaAlaI 994
3229 .....CTGATTGCGCGCGCGCGCAT 3249
994 leLeuArgAsnAspGlyLysSerPheTyrIleLeuAlaThrAspLysAsn 1010
3250 GCGCGCGCAAAAGACAGCAAGC..... 3270
1011 ThrAlaGluLysSerAspAlaAlaAsnGlyAspTrpAsnGlyLeuArgPr 1027
3271 .....GTTGCCGAACCGCGCGCGAGCGGGAATGTCGCGCATTA 3315
1027 oPheAlaIleAsnMetAlaAspGlyArgValGlyMetAsnHisGlyLeuA 1044
3316 T...GCAGCGGAGAGAGAGAAAAACGGGTGCAGCGCGGATTAAGACAG 3352
1044 snIleThrGlyGlyLeuAsnValThrGly..... 1054
3363 CGCNTTGGCAAAACAGCGCAAGGGAACCGCGCGGTACACCGCGCT 3412
1054 ..... 1054
3413 TCCTCCGCGCGCGCGCGCGCGGATTTGCGCAACCGCAGCCCA 3462
1055 .....GlyLeuAsnValThrSerGlyAsnT 1063
3463 CCGCACTCAACCCCAACCGCAGCGCATCTGATNAGCCGTTATGCCAA 3512
1063 hSerLeuGlyAsnIleSerSer ArgValValAlaArgTyrPargGlyAl 1079
3513 TAGCGGTGAGTGAATTTCCGCGCAAGCTCAACAGCGTTTCGCGGAC 3552
1079 aSerGlyTrpAlaAspAsnSerAspThrMetLysSerLysIleThrPheM 1096

```

```

3563 AGGACGAATTTGACCGCGTGTGGCCGAAGACCGCGCAACCGTGTGG 3612
1096 eTrAlaSpHISGlyAspLeuSerAsnSerAsp..... 1106
3613 ACAGCNGCATCCGAGACCAACACTACCTTGCAGATTTCCGCC 3662
1107 ...SerTyrTrpIleValGlyAlaTyrSerAsnTyrGlySerAlaG 1122
3663 CTACGCGCAAAACCGACCTGCGCAATCGTATGAGAAAACCTG 3712
1122 yTrArgGlnThrPheGlnPheIleTyrValGlySerGlyThrAlaG 1139
3713 GCAGCGCGCGCGTGCATCTGTTTTCGCAACCGCAGCAACAAAC 3762
1139 Ly...TrpArgAspGlyIleIleArgIleArgGly..... 1149
3763 TTGACAGAGCGCATCGCAACTCGCGAGCGCTTGCCAGCGCGCTTT 3812
1150 ...AspAsnAlaAsnGlyGlnAlaArgTyrParg.....Ph 1161
3813 CGGCAATACGCG.....A 3826
1161 eTrMetAspGlyThrLeuAspCysProGlyLysValLeuLeuProGlnT 1178
3827 TCAGCAGGTTTCGACATCGGCATCAGCAGCGCGGTTTTCAGACGCG 3876
1178 hrcGlyAlaPheGlyValAsnThrSerAsnGlyLeuGlyLysAsnTyr 1194
3877 ANCTNTACAGCGCATCGAGCAAAATCCGCGCGCGGTGCTGATTA 3926
1195 ThrPheLysSerAspThrGly...IleLysGlnAsn..... 1206
3927 CGCATTCAGCAGCATACCGCGCGGTTTCGCGGATTCGCAATCAGAC 3976
1207 .....GlyAspGlyLeuLeuAspI 1213
3977 CGTACATCGCGCGCAACCGCTATTTC...GTCCAAAACCGGAT...TAC 4020
1213 leTyrAlaAsnSerValGlnValPheArgPheGlnAsnGlyAspLeuTyr 1229
4021 CGGTACGAAACGTCATATCGCCACCGCGGCTGTGCTTCACCGGTA 4070
1230 SerTyrLysAsnIleAsn.....AlaProAsnValTyr 1240
4071 C...CGNCGCGCATTAAGCAGATTTATTCATCAACCGCGCAACAC 4116
1240 rIleArgSerAspIleArgLeuLysSerAsnPheLysProIleGlnAsn 1256

```

seq_name: SwissProt_40:VGIX_HSVEB

```

seq_documentation_block:
ID VGIX_HSVEB STANDARD: PRT; 797 AA.
AC P28968:
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Glycoprotein X precursor.
GN 71.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1."
RL Virology 189:304-316(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```


16 / I n r l n r A l a i h r T h r . T h r V a l P r o T h r T h r A l a S e r T h r T h r T h r a s p 183

Saccharomyces cerevisiae (Baker's yeast).

FIG2 OR YCR089W OR YCR89W OR YCR1102.

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes
OC Eukaryota; Fungi; Ascomycota (Baker's yeast),
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos

NCBI_TaxID=4932;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92397594; PubMed=1523889;
 RA Wilson C., Grisanti P., Frontali L.,
 RT "The complete sequence of a 6146 bp fragment of Saccharomyces
 cerevisiae chromosome III contains two new open reading frames."
 RL Yeast 8:569-575(1992).
 CC
 CC -1- FUNCTION: REQUIRED FOR EFFICIENT MATING.
 CC -1- INDUCTION: BY MATING PHEROMONES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X59720; CAA42254.1; -
 DR PIR: S19504; S19504.
 DR PIR: S25345; S25345.
 DR SGD: S0000685; FIG2.
 DR
 SQ SEQUENCE 1609 AA; 166049 MW; 7D66AD7F85A7B852 CRC64;

alignment_scores:
 Quality: 183.00 Length: 1228
 Ratio: 0.345 Gaps: 52
 Percent Similarity: 43.241 Percent Identity: 18.241

alignment_block:
 US-09-303-518D-651 x FIG2_YEAST ..

Align seg 1/1 to: FIG2_YEAST from: 1 to: 1609

892 AAAGATTGTTCTACGATGACATTTACAGGCGGATACACATACCGTCTN 941
 :::::||||| :::::||||| :::::|||||
 408 GlnAspIrrSerSerSerSerSerLeuProLeuSerGlnThrThrProValVa 424
 942 TTTTGAACCGCGCGAGTACGACATTTTCCTTACATCCACACACACG 991
 :::::||||| :::::||||| :::::|||||
 424 LileAsn.....ThrThrAsnThrGlnG 432
 992 GTACGGGAGGTACAGAAACCAACGAAAGTTCATCCAAAGCTT 1041
 :::::||||| :::::||||| :::::|||||
 432 lYserValThrSerThrSerProAlaThrValSerThrAlaThrLys 448
 1042 AAAGTACAGACGTCGACTGTTTGACGAA.....TCTTT 1076
 :::::||||| :::::||||| :::::|||||
 449 ThrValAspGlyVal.....IleThrGluThrValThrTrpCysProLe 463
 1077 GAATGAACAGTATTAAGACACGTTTACGGCGCGGGGTGTAATAGT 1126
 :::::||||| :::::||||| :::::|||||
 463 uThrGlnThrLysSerGlnAlaIleGlyValSerSerIleSerSerV 480
 1127 ACCGTCAAGGTAAACACGCTGAAAACCTTTCTTATGATTAACGCG 1176
 :::::||||| :::::||||| :::::|||||
 480 al...ProGlnAlaSerSerPheSerGlySerSerIleLeuSerSerAsn 495
 1177 AACGGCAACTCATCTTATCAACACATCAACCAAGC..GGGGCGG 1223
 :::::||||| :::::||||| :::::|||||
 496 SerSerThrLeuAlaAlaSerAsnAsnValProGlnSerThrAlaSerG 512
 1224 TTTGATTTTGAAGGTATTTTACGGTCTG.....CCTGAACAACAGC 1267
 :::::||||| :::::||||| :::::|||||
 512 ySerSerGlnTyrGlnAspIrrSerSerSerSerLeuProLeuSerGlnT 529
 1268 AAAGCTGG.....CAAGCGCGGGGGCTTCAT 1293
 :::::||||| :::::||||| :::::|||||
 529 hrrThrTrpValValIleAsnThrThrAsnThrGlnGlySerValThrSer 545
 1294 ATCAGTGAAGAC.....AGTACCGTTACTTGAAAGTAACGCGCGT 1334

546 ThrThrSerProAlaThrValSerThrAlaThrLysThrValAspGlyVa 562
 :::::||||| :::::||||| :::::|||||
 1335 GGCAAGACAGCCGCTGTCGCAAAATCGCAAAAGGACCGCTCAGCTTCAG 1384
 :::::||||| :::::||||| :::::|||||
 562 lIleThrGluThrValThrTrpCysProLeuThrGlnThrLysSerGln 579
 1385 CCAAGGGGAAACCAAGGCTGATCAGCGTGGCGACGCTACAGTCATT 1434
 :::::||||| :::::||||| :::::|||||
 579 lAlaIleGlyIleSerSerSerThrIleSer..... 588
 1435 TTGATTCAGCGGACGATTAAGGCAAAAAACACGCTTTAGTGAAT 1484
 :::::||||| :::::||||| :::::|||||
 589AlaThrGlnThrSerLysProSerSerIleLeuThrIle 601
 1485 CGGCTTGNTTCAGCGGAGGGGTACGTCGACATGATGCGATATACGT 1534
 :::::||||| :::::||||| :::::|||||
 601 uGlyIle.....SerThrIleGlnLeuSerAspAlaThr.... 612
 1535 TCACCCCGACAAACTCTATTTCGCTTCGCGGGGACGCTTGATTTA 1584
 :::::||||| :::::||||| :::::|||||
 613PheLysGly..... 615
 1585 AACGGGCAATTGCTTTCGTTCCACCGATTCGAAATACCGATGAAGGCG 1634
 615 615
 1635 GATGATTGNCATCATTAATGCCACACACATCCACGCTTACATTACAG 1684
 :::::||||| :::::||||| :::::|||||
 616ThrGluThrIleAsnThrHisLeuMetThrG 626
 1685 GAATGAAGATTTACACACCGAGTGTGAAGATATCATATAGACTTAAT 1734
 :::::||||| :::::||||| :::::|||||
 626 lSerThrSerIleThrGluProThr..... 634
 1735 TACACCAAGAAATTCCTACAAAGTGTGGCGAGAAAGATACGAC 1784
 634 634
 1785 CAAAGCAAGGCGGCGCTCAACCTGTTTACCAGCCGCGCAGACAGCC 1834
 :::::||||| :::::||||| :::::|||||
 635TyrPheSerGlyThrSerAspS 642
 1835 GCACCCNGCTGCTTCCGGCGGACAAATTTAAAGGCAATCAGCGCA 1884
 :::::||||| :::::||||| :::::|||||
 642 ePheTyrLeuCysThrSerGluValAsnLeuAlaSerSerLeuSer... 657
 1885 ACAAAAGCAACTGTTTTCAGCGGACGACGACGACGCGCTACAA 1934
 :::::||||| :::::||||| :::::|||||
 658SerTyrPr 660
 1935 TCATTTAGAGAGCGGGGTGTCAAAAATGAGAGTATCCACAGAGCAAA 1984
 :::::||||| :::::||||| :::::|||||
 660 oAsnPheSerSer..... 665
 1985 TCGTGTGGACACAGATGATCMACGACGTTAAAGCGGAAATTTTC 2034
 665 665
 2035 CATATTCAAGCGGCGGAGCGGTGATTCGCCCAATGTTGCCAAAGTGA 2084
 :::::||||| :::::||||| :::::|||||
 666GluGlySerThrAlaThrIleThrAsnSerThrValThrPheG 680
 2085 AGGC.....GATTGCAATTGGACATCAACGCCCAAGCAG 2119
 :::::||||| :::::||||| :::::|||||
 680 ySerThrSerLysTyrProSerThrSerValSerAsnProThrGluAla. 696
 2120 TTTTGTGTGCGACGCAATCAAGGATATACATGTACAGCTTCGCGAC 2169
 :::::||||| :::::||||| :::::|||||
 697SerGlnHisValSerSerSerValAsnSerIleuThrAsp 709
 2170 TGCACNGGTCTGACAAATTTGTGCAANAANCAATTACGACGATTAAGT 2219
 :::::||||| :::::||||| :::::|||||

```
710 pheThr.....SerAsnSerThrGluThrIle.....AlaVal 720
2220 GATTGGTTCATTGACTAGACACGACNTNAGCGGCANTGNAGCTNCCN 2265
720 IILSerAsnIleHisLysThr..... 727
2270 ATNAGCNTNNTNNAANCTNCNCGGCTGCTNCACTNANGCAATCTT 2319
727 ..... 727
2320 AGTGAATAGCGGATACAGCTTATACAGTACGCCAACGCCACCAAAA 2369
728 SerSerAsnLysAspTyrSerLeuThrThrThrGlnLeuLysThr...Se 743
2370 CGGCAACCTTACGCTCGTGGGCAATGCCCAAGCAATTTATCAAGCCA 2419
743 rGlyLysGlnThrLeuValLeuSerThrValThrThr.....T 756
2420 CATTAACGGCAACNCATCGGNTTCGGGCAAT.....GCTTCATTT 2460
756 hrValAsnGlyAlaAlaThrGluThrThrThrThrPcysProAlaSer 772
2461 AATCAACCAACACAGCCGCACAAAAGCGAGCTGACGCTTCCGACAA 2510
773 IleAlaTyrThrThrSerIleSerTyrLysThrLeuValLeuThrThr 789
2511 CGCTAAGCAACAGTAAGCATTCGCGACCTCACGGCAATGTCTCCCTAG 2560
789 uVal.....CysSerHisSerGluCysThrProThrVal..... 800
2561 CGGATAGGCGATATTCATTTTGAAAACAGCCGCTTTACCGCAACCTC 2610
801 .....IleThrSerValThr 805
2611 AGCGGACAGAGANACAGATTAACATTAAGACAGACGAGATGAGAGCT 2660
806 AlaThrSerSerThrIleProLeuLeuSerThrSerSerThrValIle 822
2661 GCCGTCAAGCAGCAATTAAGCAATTTAAACCTTGACACGCCACCACTTA 2710
822 uSerSerThrValSerGluGlyAlaLysAsnProAlaIleSerGluValT 839
2711 CACATTCGCGCTATCGCCACAGATGCTGACGGCGCAACCGCGCAGN 2760
839 hrIleAsnThr.....Gln 843
2761 GTGTCAAGACAGCCGCCGCGCTTCGCGCGCTTCCTATTATCCGTTAC 2810
844 ValSerAlaThrSerGluAlaThrSerThrSerThrGlnValSerAlaThr 860
2811 ACCGCCAATCTGGTAGAATCCGCTTCAACACGCTGACGCTAAACGCA 2860
860 rSerAlaThrAlaThrAlaSerGlnSerThrThrSer..... 873
2861 AATGACACNGTCAAGAACATTCGCTTATGTCGGAACCTTCGCTAC 2910
873 ..... 873
2911 CGAAGCACAATTAAGCTGCGGGAAGTTCGAAGGNACTTACACCTT 2960
874 .....GlnValSerThrAlaSerGlnThrIleSerThrIle 885
2961 GCGGCGTAACAAT.....ACCGGCAAC.....GAACCGGTAA 2992
885 uGlyThrGlnAsnPherThrThrThrGlySerLeuLeuPheProAlaLeuS 902
2993 GCGTCGATCAATG.....ACGCTAGTGAAGGAAGAACACA..... 3031
902 erThrGlnMetIleAsnThrThrValVal.SerArgLysThrLeuIleI 918
3032 .....AACGCTGTCGGAACACCTTAATTTAC 3059
918 eSerThrGlnValCysSerHisSerLysCysValProThrValIleThrG 935

3060 CTTGCAAAAGCAACACGTCGATGCCGCGCGCTGAC..... 3094
935 luValValThrSerLysGlyThrProSerAsnGlnHisSerSerGlnThr 951
3095 .....GTTAC 3099
952 LeuGlnThrGluAlaValGluValThrLeuSerSerHisGlnThrValThr 968
3100 CAACCTCATCCGCAAAAGACGGGAGTTCCGCTCCATTA..... 3136
968 rMetSerThrGluValCysSerAsnSerIleCysThrProThrValIleT 985
3137 .....ATCCGCTCAAGAACACAGAGCTTTCG.....ACAACTCGCA 3175
985 hrSerValGlnMetArgSerThrProPheProIleuThrhrSerSerThr 1001
3176 AGGCAGAGCCCAAAACAGAGCGGAAAAAGACAGCGCAAAAGCTTGAC 3225
1002 SerSerSerSerLeuAlaSerThrLysLysSerSerLeuAla..... 1016
3226 GCGCTATTTCGGCGCGCGCATGCCGCGAAAGACAGAAAGCGTTTC 3275
1017 .....SerSerGlnMetSerT 1022
3276 CGAACCGCGCGCGCGCGAGCGGAGGAAATGCGCATTAATGACAGCGG 3325
1022 hrPheSerValSerThrGlnSerLeuProLeuAlaPheThrCysSerGln 1038
3326 AGGAGAGAAAAAGGGTGCAGCGCGGATTAAGACAGCCGCTTGGCGGAA 3375
1039 LysArgSerThrThr.....SerValSerGlnThrPheSer 1050
3376 CACGCGGAAGCGGAAACCGCGCGGTACACCGCTTCCCGCGCGCGG 3425
1050 nThr..... 1051
3426 CNGCGCGCGCGCGGATTCGCGCAACCGCAAGCCCAACCGCAACTCAAC 3475
1052 .....ValLeuThrAsnThrIleMetSerSerSerSerAsn 1063
3476 CCCAACCGAGCGCGACATGATNAGCGCTTATGCCAATAGCGGTTGAGT 3525
1064 .....ValIleSerThr 1067
3526 GAATTTTCCGCCACGCTCAACAGCTTTTGGCGTAACAGAGCAATGGA 3575
1067 rAsnGluLysProSerSerThrThr...SerProTyrAsnPheSerSerG 1083
3576 CCGCGTGTTCGCGAAGACCGCGCGCAACGCTTGGACACACGNCATCC 3625
1083 lYTySerLeuProSerSerSerThrPro.....SerGlnTyrSerLeu 1097
3626 GGNACACCAACACTACGCTTCGCAAGATTTC..... 3658
1098 SerThrAlaThrThrThrIleAsnGlyIleLysThrValTyrThrThrT 1114
3659 .GCGCTACCGCAACAAACGACCTGCGCC.....A 3689
1114 pcysProLeuAlaGluLysSerThrValAlaIleSerSerGlnSerSerA 1131
3690 AATCGGTATGCAGAAAAACCTGCGAGCGGCGCGCTCGGCAATCCGTTT 3739
1131 rGSerValAspArgPheValSerSerSerLysProSerSerSerLeuSer 1147
3740 CGCACACCGGACGCAAAACACTTCAGACAGCGCATCGCAACTCGCA 3789
1148 GlnThrSerIleGlnTyrThrLeuSer...ThrAlaThrThrThrIleSe 1163
3790 GGGCTTGGCCGACG.....GCGCGCTTTCGGGCAATTAC. 3823
1163 rGlyLeuLysThrValTyrThrThrThrPcysProLeuThrSerLysSerT 1180
```



```

2800 .....TTATCCGTTACACGCCCACTTCGTAGAT 2830
1378 pValAspAlaGlyLeuSerHisSerAsnAspProGlySerSerIleThr 1395
2831 CCCGTTTCAC.....ACGCTGACGGTAAACGCCAATTAAGACNGT 2871
1395 erLysLeuSerIleValGlyThrProArgTyrAlaGlyValIleGlu 1411
2872 CAAGAACATTCGCTTATGTCGGAACCTTCGGCTACCGAACGCGAA 2921
1412 ..LysLeuIleAlaGlyVal.....AsnLysValAlaLaspAlaThr 1425
2922 ATTGAAGTCGGCGAAGTCCGAGAGNACTTACACCTTCGCGCTCA... 2968
1425 hrAspLysTyrAsnSerValAlaArgArgLeuAspProGlnGlnAspThr 1441
2969 .....ACAAATACCGCGC 2979
1442 ThrGlyAlaValSerPheSerIleAlaGluGlyLysValThrLeuProAl 1458
2980 AAGC.....AACCGTAAGCCTCGATCAATTGACGGTAGTGA 3017
1458 atHProAlaGlyGlyLysPro.....G 1466
3018 AGGAAAGACAAACACCGCTGTCGAAACCTTAATTCACCTCGCAA 3067
1466 lngIyProLeuTyrAspArgGlyAlaArgThrValGlyGlyAlaValLys 1482
3068 ACGAACACGTGATGCCGCG.....CG 3090
1483 AspSerIleThrGlyProAlaGlyArgGlnGlyHisLeuLysValAsnAl 1499
3091 TGGCGTTACCAATCCGCAAGACGGCGAGTCCGCTCGATATACC 3140
1499 aspValAlaValAsnAsnAlaValGlyGlnGlnSerAla.....IleA 1514
3141 GGTCAAGACAGAGCTTCCGCAACACTGCGACGACGACGACCAAA 3190
1514 lngIyLysAsn.GlyValAlaLeuGlnVal...GlyGlyGlnThrGlnLe 1529
3191 AACAGCGCGAAGACACGCGCAAGACCTTGACGCGCTATTCGGCC 3240
1529 urHcIyGlyLys...IleArgSerGlnGlnGlyValGlnLeuGlyG 1545
3241 GGGCGCGATGCCGCGAAGACGAAAGCGTCCGCAACGCGCGCGCN 3290
1545 LysSerGlnValSerGlnGlnAspValAsn..... 1554
3291 GGCAGCGCGGGAATAATGTCGATTAATGACGCGAGAGAGAGAAAAAC 3340
1555 GlyIleArg.....TyrGlnGlyGlyArgValAspAl 1566
3341 GGGTGCA.....GGCGATTAAGACAGCGCNTTGGCGAAACACGCGAA 3384
1566 alaIaIaThrValGlyGlyLeuGlyGlyAlaValGlnSerValA 1583
3385 GCGGAAC 3392
1583 laGlyAsn 1585

```

seq_name: SwissProt_40:P3P_LACLC

seq_documentation_block:

ID P3P_LACLC STANDARD; PRT; 1902 AA.

AC P15292;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE P11-type proteinase precursor (EC 3.4.21.96) (Lactocoeplin) (Cell wall-associated serine proteinase).

DE associated serine proteinase).

OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).

```

OG Plasmid.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus1585;
OX NCBI_TaxID=1359;
RN
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
RX STRAIN=SK11;
RX MEDLINE=89340435; PubMed=2760036;
RA Vos P., Simons G., Siezen R.J., de Vos W.M.;
RT "Primary structure and organization of the gene for a procaryotic,
RT cell envelope-located serine proteinase.";
RL J. Biol. Chem. 264:13579-13585(1989).
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some substrate preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC Pro in the P2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyse hemoglobin and oxidized
CC insulin B-chain.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04962; AAA0353.1; ALT_SEQ.
DR PIR: A32634; A32634.
DR HSRP: P00782; 2SBR.
DR MEROPS: S08.019; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR003137; PA.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 3.
DR PRINTS: PR00723; SUBTILISTM.
DR PROSITE: PS00136; SUBTILISTM_ASP; 1.
DR PROSITE: PS00137; SUBTILISTM_HIS; 1.
DR PROSITE: PS00138; SUBTILISTM_SER; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
KW Hydrolase. Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane.
KW SIGNAL.
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
FT PROTEINS.
SQ SEQUENCE 1902 AA; 200550 MW; 87CECBA9345F9D3 CRC64;

```

alignment_scores:

Quality:	Length:
Ratio: 0.265	1558
Percent Similarity: 44.031	Gaps: 74
Percent Identity: 18.678	

alignment_block:

US-09-303-518d-651 x P3P_LACLC ..

Align seg 1/1 to: P3P_LACLC from: 1 to: 1902

[illegible]

663 ValGluMetAsnThrAlaGlnProIleAsnSpleAsnTyTrpAsnAspVal 679
864 |||::: ||| ::: 864
679 IleValSerProArgTrpGlnGlyAlaGlyLeuValAspAlaLysAla 696
865 |||::: |||::: 876
696 IleAlaSerAlaLeuGlnLysAsnProSerThrValAlaIleAlaGlnGly 712
877 TTCACAGCTGATGCG...AAGATTGGTTCTACAGATGAC..... 912
713 TyrProAlaValAlaGlnLeuLysAspPheThrSerThrAspLysThrPhe 729
913ATTTCACAGAGCGCATACATACCGCTGTTTTCACACCGCGCA 955
729 SLeuThrPheThrAsnAlaGlyThrThiHisGlnLeuThrTyrGlnMetAsp 746
956 GTAACGGA.....CATTTTCCTTTACATCCACACACAC 990
746 eTrasnThrAspThrAsnAlaValTyrThrSerAlaThrAspProAsnSer 762
991 GGTACG.....GGTACGGTAACAGAAACCAACGA 1019
763 GlyValLeuTyTrpAspLysLysIleAspGlyAlaAlaIleLysAlaGlySe 779
1020 AAAGGTTTCCATCAAAAGCTTAAAGTTCACAGACAGTCCGATGTTTACAG 1069
779 rAsnIleThrValProIleGlyLysThrAlaGlnIleLys...PheThrL 795
1070 AATCTTTGAATGAACACTGATTAAGAACCAAGTTTACGCGGACGGGGGTGT 1119
795 euserLeuProLysSerPheAspGlnGlnIlePheValGlnGlyPheLeu 811
1120 AATCACTACCGTCCAAAGGTTAAACACGGTGAAACCTTCT..... 1161
812 AsnPhe.....LysGlySerAspLysSerArgLeuAsnLeuProTy 825
1162TTTACGATTTACGCGACAGCGCAACCTCATCTTACAAACA 1201
825 rMetGlyPhePheGlyAspTPTrpAsnAspGlyLysIleValAspSerLeuA 842
1202 AC..... 1203
842 singlyLeuThrTyTrSerProIleAlaGlyLysAlaPheGlyThrValProLeu 858
1204 ..ATCAACCAAGCGCGCGCGTTTGAATTTTGAAGGTGATTTTACGGT 1250
859 LeuLysAsnLysAsnThrGlyThrGlnTyrTy...GlyGlyMetValTh 874
1251 CTCGGCTGAAACACACGAACAGCTGGCAAGCGCGCGCGCTTATATCACTG 1300
874 rAspAlaAspGlyAsnLysThrValAspAspGlnAlaIleAlaPheSerS 891
1301 AAGACAGTACCGCTTACTGGAAGATGAACGGCGTGCGCAACAGCACCGCTG 1350
891 eTrAspLys.....AsnAlaLeuTyTrpAsnAspLysSer 901
1351 TCCAA.....ATCGCAAAAGCGACGCTGCACGTTCAAGCCAAAGGGGA 1394
902 MetLysTyTyLeuLeuArgAsnIleSerAsnValGlnValAsp..... 916
1395 AAACCAAGGCTGATCAGCGTGGCGGACGAGTACAGTCAATTTGGATCAGC 1444
917IleLeuAsp.... 919
1445 AGCAGACGATAAAGCAAAACACAGCTTAGTGAATCGCTTGTC 1494
920GlyGlnGlyAsnLysValThrThrLeuSerSerSer 931
1495 AGCGCAGGGGTTACGTCCAACTGANTCCGATATATCAG..... 1533


```

3168 ACTCGGCAAGCAGAAAGCCAAAAAGCGGCAAAAGCAACAGCGCAAA 3217
1424 ThrGlyThrAla.....ThraspsnaIaGInt 1433
3218 GCGTTGACCGCGCTGATGCGCGCGCGGATGCCGCAAAAGCAAGAA 3267
1433 yrlEu...SerIleuSerIleuSerIleuSerIleuSerIleu 1448
3268 .....AGCGTTGCCGCAAC 3281
1449 AspIleuAsnIleuSerIleuSerIleuSerIleuSerIleuSerIleu 1465
3282 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3331
1465 oValIleuSerIleuSerIleuSerIleuSerIleuSerIleuSerIleu 1477
3332 AGAAAAAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3381
1478 .....AlaValThrAspSerIleuSerIleuSerIleuSerIleu 1491
3382 GAAGCGGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3431
1492 ThrValIleuSerIleuSerIleuSerIleuSerIleuSerIleu 1497
3432 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3478
1498 LysIleuSerIleuSerIleuSerIleuSerIleuSerIleuSerIleu 1514
3479 AACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3525
1514 IaGIntThrValIleuSerIleuSerIleuSerIleuSerIleuSerIleu 1530
3526 GAATTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3575
1531 GlnIleuSerIleuSerIleuSerIleuSerIleuSerIleuSerIleu 1580
3576 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3625
1541 .....GlnAspValProAlaIaGIntThrIleuSerIleu 1553
3626 GGNACCAACCAACACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3675
1553 snGIntThrIleuSerIleuSerIleuSerIleuSerIleuSerIleu 1580
3676 ACCGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3725
1587 userProAlaValIleuSerIleuSerIleuSerIleuSerIleu 1580
3726 CGGACCTGCTTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3775
1581 .....AspAspProAlaIleuSerIleu 1586
3776 TCGGCAACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3825
1587 GlnAlaIleuSerIleuSerIleuSerIleuSerIleuSerIleuSerIleu 1603
3826 ATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3875
1603 userAlaSerIleuSerIleuSerIleuSerIleuSerIleuSerIleu 1620
3876 CANTC.....TNTCAGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 3919
1620 IaThrGlnIleuSerIleuSerIleuSerIleuSerIleuSerIleuSerIleu 1636
3920 TGCATTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3945
1637 SerIleuThrGlnIleuSerIleuSerIleuSerIleuSerIleuSerIleu 1653
3946 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3995
1653 AlaIleuSerIleuSerIleuSerIleuSerIleuSerIleuSerIleu 1667
3996 CTATTGTCGCAAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4045

```

```

1667 InleuGlnSerValIleuSerIleuSerIleuSerIleuSerIleu 1683
4046 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4095
1684 AsproSerThrGlnIleuSerIleuSerIleuSerIleuSerIleu 1699
4096 TCATTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4117
1699 IaIaGIntThrValIleuSerIleuSerIleuSerIleuSerIleu 1706

```

seq_name: SwissProt_40:FNAB_BORPE

seq_documentation_block:

```

ID FNAB_BORPE STANDARD; PRT; 3591 AA.
AC P12255;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Filamentous hemagglutinin.
GN FNAB.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90355839; PubMed=2388559;
RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Genetic characterization of Bordetella pertussis filamentous
RT haemagglutinin: a protein processed from an unusually large
RT precursor."
RL Mol. Microbiol. 4:787-800(1990).
RN [2]
RP SEQUENCE OF 1-3261 FROM N.A.
RX MEDLINE=89202384; PubMed=2539596;
RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Filamentous hemagglutinin of Bordetella pertussis: nucleotide
RT sequence and crucial role in adherence."
RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).
CC -1- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND
CC INFECTION.
CC -1- SUBCELLULAR LOCATION: SURFACE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M60351; AAA22974.1; -
DR EMBL; M60351; AAA22975.1; ALT_INIT.
DR EMBL; M60351; AAA22976.1; ALT_INIT.
KW Antigen; Hemagglutinin.
SO SEQUENCE 3591 AA; 367420 MW; EF741BB30D6E5138 CRC64;

```

alignment_scores:

```

Quality: 180.50 Length: 1427
Ratio: 0.305 Gaps: 71
Percent Similarity: 41.416 Percent Identity: 19.131

```

alignment_block:

```

US-09-303-518d-651 x FNAB_BORPE
Align seg 1/1 to: FNAB_BORPE from: 1 to: 3591

```

```

202 GCGAAGATATGAGGTATACCAAAAAAGCGGAG..... 237
2290 AlaGlnIleuGlnValSerIleuSerIleuSerIleuSerIleuSerIleu 2306

```


2738 uGlyAlaLysValGlyAla.....GlyGlyT 2747
1859 CAATTTAAAGCGCAATCATCGCAAAACGCAAACTGTTTTCAGC 1908
2747 YrGluAlaGlyPheSerLeuGlySerGlyLeuGluAlaHisAla 2763
1909 GGCAGACCGCAGCAGCGCCTTACATATTAGAAACGGGTGCTAAA 1958
2764 GlyArg..... 2765
1959 AATGGAAGTATCCACAGAGAAATCGT..... 1989
2766GlyMetThrAlaGlyAlaGluValLysValGlyTyrArgAla 2780
1990TGGACAACAGACTGATCAGTCAGTCATTAAGCGGAAATTC 2034
2780 erHisGluGlnSerSerGluThrGluGlySerTyrArgAsnAlaSerLeu 2796
2035 CATATTTCAGGCGGCGCAG..... 2052
2797 AsnPheGlyGlyGlySerValGluAlaGlyAsnValLeuAspIleGly 2813
2053 .GCGGTGATTCCCGCAAT..... 2070
2813 yAlaAspIleAsnArgAsnArgTyrGlyGlyAlaAlaLysGlyAsnAlaG 2830
2071GTTGCCAAAGTGAAGCGGATTCG 2094
2830 LyThrGluGluAlaLeuArgMetArgAlaLysValGluSerThrLys 2846
2095 CATTTGACCAATCACGCCCAAGAGTTTGGTGCACCGCATCAAG 2144
2847 TyrValSerGluGlnThrSerGlnSerSerGlyTyrPheValGluVal 2863
2145 CCATACATCTGTACGTTGCGAGCTG.....ACNGTCTGA 2182
2863 aSerThrAlaSerAlaArgSerSerLeuThrAlaAlaThrArgLeuG 2880
2183 CAATTTGTGCAANAAACATTACGACGATAA..... 2217
2880 LysPheSerValAlaGlnAsnValGluAspGlyArgGluIleArgGly 2896
2218 GTGATGCTTATTGACTAGACAGACNTNAGCGGANTGNNCTNNC 2267
2897 LeuMetAlaAlaGlnValAlaAlaGlnAlaThrGlnLeuValThrAla 2913
2268 CNAATNAGCTNNNTNAAANCTNCGGCTNCACTNCACTNAGCAATC 2317
2913 P.....ThrAlaAlaValAlaLeuSerAlaGly 2923
2318 TTAGTGCATATGCGATACAGTTATACAGTCAACGCAACGCCCAA 2367
2923 LeSerAlaAspPheAspSerHis.....SerArgSerThrSerGln 2937
2368 AAC.....GCCAACCTTACG..... 2382
2938 AsnThrGlnTyrLeuGlyGlyAsnLeuSerIleGluAlaThrGlnGly 2954
2383CTCGTGGGC..... 2391
2954 PAlaThrLeuValGlyAlaLysPheGlyGlyAspGlnValSerLeu 2971
2971 yAlaAlaLysSerValAsnLeuMetAlaAlaGlnSerThrPheGlnSer 2987
2416 GCCCATTTAAAGCGCAACNCATCGGNTTCGGCAATGCTTCAATTACT 2465
2988 TyrSerGlnSerHisAsnPheHisAlaSerAlaAspAla.....AsnLe 3002
2466 AAGCAACAACCGCGCACAAACGCGAGTCGACGTCGACCAACGCTA 2515
3002 uGlyAlaAsnAlaValGln...GlyAlaValGlyLeuGlyLeuThrAlaG 3018
2516 AGGCAACGTAAAGCCATTCGCGACTCAAC..... 2544
3018 LyMetGlyThrSerHisGlnIleThrAsnGluThrGlyLysThrTyrAla 3034
2545GGCAATGCTCCCTA...GCCGATTAAGCGAGT 2573
3035 GlyThrSerValAspAlaAlaAsnValSerIleAspAlaGlyLysAsp 3051
2574 ATTCCATTTTGAACACGCCGCTTACCGGACACTACGCGGACAGAG 2623
3051 u.....AsnLeuSerGlySerArgV 3058
2624 ANACCATTTACCTTA..... 2640
3058 aLArgGlyLysHisValValLeuAspValGluGlyAspIleAsnAlaThr 3074
2641 .AAGACAGCGA.....TGAGCGCT 2660
3075 SerLysGlnAspGluArgAsnTyrAsnSerSerGlyGlyTyrPheAsp 3091
2661 GCCGTACGACGAGAAATTAGCAATTTAACTT..... 2694
3091 aSerAlaGlyValAlaIleGlnAsnArgThrLeuValAlaProValGly 3108
2695GACACGCCACCATTAACACTC 2715
3108 erAlaGlyPheAsnPheAsnThrGlnHisAspAsnSerArgLeuThrAsn 3124
2716 AATTCCGCTATTCGCCAGATGCTGACGCGGCAACCGGAGNGTGT 2765
3125 AspGlyAlaAlaGlyValAlaValaSerAspGlyLeuThrGlyHisVal 3140
2766 AGACACGCCGCGCGCGCTTCCGCTTATATCCGTTACACGCC 2815
3141LysGlyAspAlaAsnLeuThrGly 3149
2816 CAATTCGTGAGAA...TCCCGTTTCAACAGCTGACGTAACGCAAA 2862
3149 LeThrIleAlaAspLeuSerGlyLysGlyAsnLeuLysValAspGlyAla 3165
2863 TTGACNGTCAAGAAACATTCCGCTTATGTCGGAACCTTCGCGCTAC 2912
3166 ValAsnAlaGln.....AsnLeuLysAspTyr 3175
2913 AAGCGCAAAATTAGCTGCGGAAAGTCCGAAGNACTTACACTTGG 2962
3175 g...AspLys.....AspGlySerGlyGlyLeuAsnValG 3187
2963 CGGTCAACATACCGGCAACGACCGTAAAGCTCGATCAATTGACGTA 3012
3187 LyLeSerSerThrThrLeuAlaProThrValGlyValAlaIleGlyArg 3203
3013 GTGGAAAGGGAAGACAAACCGCTGTCGGAACCTTAATTTCACCT 3062
3204 ValAlaGlyGluAspTyrGlnAlaGlnArgAlaThrIleAspValG 3220
3063 GCAAAAGAAACAGCTGATCGCGCGCT.....GAGCGT 3097
3220 LysIleThrLysAsp.....ProAlaArgLeuGlnValAlaGlyGlyVal 3234
3098 ACCAATCATCCGCAAGAGCGGAGTTCGCTGATATCCGTCANA 3147
3235 LysGlyThrLeuAsnGlnAspAlaAlaGlnAlaThrValValGln...Ar 3250
3148 GAACAAG.....AGCTTTCCGACAACTGGCAA 3176
3250 gAsnLysHisThrPheAlaGlyGlyLysSerGluPheSerValAlaGlyLys 3267
3177 GGCAGAAGCCAAACAGCGGAAAGAAAGACAAAGCGCAACGCTTGAAG 3226
3267 erLeuLysLysLysAsnGln..... 3273


```
762 AGTGGCGGACGACAGCGGTTGCCAATGTTATTATTATGCAAAACA. 810
      :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
153 yGIyAlaIleTySerSergIngluIleuLeuPheThr...AspLeuThrG 169
      :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
811 .....AACATTAATGGCTGCTCAACGAGATTTTACAAACC 846
      :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
169 IyLeuThrIleGInglYAsnLeuSergInleuSergIy..... 181
      :||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
847 GGTACCCCTTATCCGCGAGGGAAGAAAGGTTTCCAGCTGATACGCAAGA 896
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
182 gIyGIyIlePheGlySerThrIleSerPheSergIyIleAsnGlnAl 198
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
897 TTGGTTCTACGATGACATTTACAGAGCGATACACATACCGCTGTTTGTG 946
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
198 aThrPhe.....SerSerAsnThrAlaGluVal 208
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
947 AACGGCCAGTAAGGACATTTTCTTACATCCAAACAACAGGTACG 996
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
208 aProGluGlu.....ThrThrProAsnProAsnPro 218
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
997 GGTACGGTAACAGAAACCAAGAAAGGTTTCCATCCAAAGCTTAAGT 1046
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
219 gLyThrGlnThrThrThrThrSergIn.....ProSerProThrSeryVa 233
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1047 ACAGACAGTCCGACTGTTTGACGATCTTGAATGAAACTGATTAAGAAC 1096
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
233 IGIleuThrPheThrTySerSerSergThrGlnAlaAsn..... 246
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1097 CAGTTTACGGGCGAGGGGTGTTTATCATACCTCCACAGGTTAAACAC 1146
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
247 .....GlyAsnGlyAlaAspSergInThrProSerHisLysPro 259
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1147 GGTGAAACCTTTCTTTATGATTTACGACAGCGCAACTCATCTATAC 1196
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
260 gLySergIyGlyAlaIle.....TyAlaThrGlyAspLeuThrIleSe 274
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1197 A.....AACACATCAACCAAGCGCGG 1219
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
274 rAspSergIngluIleValPheSerValAsnLysAlaSerLysAsp 291
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1220 GCGGTTTGAT.....TTTGAAGGTGATTTTACGCTC 1251
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
291 IyAlaIlePheAlaGluLysAsnValSerPheGluAsnIleThrThrLeu 307
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1252 TCGGCTGAAAAACAAGAAAGCGTGCAA.....GGCGGGGCGTTGCATATCAG 1298
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
308 LysAlaIleAsnAsnGlyAlaGluGlyGlyGlyIleTyAlaIle 324
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1299 TGAAGACAGTACCGTCTTGAAGAGTAAAGCGCGTGCAAGACGCGCC 1348
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
324 rGIyAspLeuSergIleGlnSergLysGlnSergLeuPheAsnSerAsn 341
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1349 TGTCCAAATTCGCAAGGACGCTGCATTCAGCC..... 1386
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
341 hrSerLysGlnGlyGlyAlaLeuTyThrIleGlnGluAsnValAspPhe 357
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1387 .....AAGGGGAAACCAAGCGTCGATTCAG 1412
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
358 LysAspLeuGluGluIleArgIleLysTyAsnLysSergIyThrPheG 374
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1413 CGTGGCGGAGGTACAGTCAATTTTGATCAGAGCGACAGCATAAAGGCA 1462
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
374 urThrLysLysValThrLeuSergLeuProGluAlaGlnThr..... 387
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1463 AAAAACAGGCTTTAGTGAATGGCTGTGTCAGCGGACGGGTACGGTG 1512
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
388 ...AsnLysSergVal...ThrAlaAlaSerGlnSergIy..... 399
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1513 CAACCTGAATCCGATATCAGTCAACCCCGACAAACTATTTGCGCTT 1562
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
400 .....ProAsnThrThrProThrProThrProThrProThrAlaAla 412
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1563 TCGCGGCGGAGCTTTGGATTAAACGGGCAATGCTTTCGTTCCACCGTA 1612
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

413 .....LysGlyGlyGlyLeu..... 417
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1613 TTCAAATATCCGATGAAGGGCGGATGATTGNCNATCATTAATCCACACA 1662
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
418 .....TyThrGlnLysAsnLeuSergIle.....SerAsnIle 428
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1663 ACATCCACCGTTACCTTACAGGAAATGAAATATTATACCAACCCAGTGG 1712
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
429 ThrGlyIleIleGluIleThrAsnAsnLysAlaThrAspValGlyGly 445
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1713 T.....AAGATATCATATACCTTAAT 1735
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
445 yAlaTyValLysGlyThrLeuThrCysLysAspSeriHisAlaGluGlnP 462
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1736 ACAGCAAGAAATTTGCTTCAACAGCTTGTTGGCGGAGAAAGATACGAC 1785
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
462 he.....GlnLysAsnSerg 467
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1786 AAACGAGAGGGGGGCTCAACCTTGTTCACAGCCCGCGGAGAAAGACG 1835
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
468 GluLysLysGly..... 471
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1836 CACCCNGCTGCTTTCGCGGGAACAATTTAAACGGCAACATCAGCAAA 1885
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
472 .....GlyGlyLeuTyThrGlnAspThrIleThrIleus 483
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1886 CAAC.....GCCAACTGTTTTCAGCGGCGACACCGACCGACGCC 1929
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
483 eAsnLeuThrGlyLysThrLeuPheGln..... 492
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1930 TACAAATCATTTAGGAAGCGGTGTCAAATAATGGAAGGTATCCCAAGG 1979
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
493 .....GlnAsnThrAlaLysGluGluGly.....GlyG 502
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1980 AGAAATCGTGTGGACAAACGACTGATGCNACCGCATTTAAACGGGAA 2029
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
502 yLeuTyThrIleGlnGlyAspAspLysThrLeuThrIleThrIleGlyLeuAsp 519
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2030 ATTTCATATTT.....CAGGGCGGCGACGGCTG 2058
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
519 ePheCysLeuIleAspAsnThrSeraIleThrIleGlyGlyAlaTy 535
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2059 ATTTCGCGAATGTGCCAAATGCAAGCGCATTCNCAATTTGAGCAATCA 2108
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
536 ValThrLysGluIleSergIleThr.....TyThrSera 547
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2109 CGCCCAAGCAGTTTGTGTGTCACACCGCATCAAGCCATACATCTGTA 2158
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
547 rValGluGluPheProGlyIleThrProValHisGlyGluThrIleIle 564
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2159 CACGTTGAGCTGACGNGCTGCACAAT.....TGTGTCGAANAA 2199
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
564 eGlyAlaCysAlaThrGlyGlySergIyGlyValCysThrLysHis 580
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2200 ANCATTTACCGAGATAAGTGAATGCTTCATTCATGACTAAG.....AC 2240
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
581 LeuValLeuSeraSnuLeuGlnThrIleSeriIleSergLysAsnPheAlaSe 597
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2241 NCACNTNAGCGGCANTGTNAGNCTNNCNATNAGCNTNNTTNAANCNTN 2290
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
597 rGluAsnGlyGly..... 601
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2291 CNGGCGTNCNCACTNAANGCAATCTTATGCAAAATGGCGATACAGCT 2340
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
602 ..GlyAlaCysThrCysProAspAsnPheProAla.....ProThrAla 615
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2341 TATACAGTCAGCACACGCCAC.....CAAAACGG 2372
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
616 SerThrProSerThrAsnGlnThrAlaAlaProLysAspAspLysAspPh 632
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
2373 CAACCTTACGCTGTCGGCAATGCCCAAGCAACATTTATGACGCACAT 2422
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

632	eLeuLIleAspYrValValSerThrThrLleAspYsInLysAlaThrL	649
2423	TAAGCGGCACACNCATCGGTTTGGGCATGCT.....	2454
649	ysLysGLYAlaGLYValYrValAlaLysLysAlaLysLeuSerArgILeAsp	665
2455	TCATTTAAATCTAAGACAACACCCGCCACA.....AACGGACATCTGAC	2498
666	GLuLeuAsnILeSerAspSrnAlaInGLInGLuThrGLYGLYGLYPro	682
2499	GCATTCCGACAAACGCTAAGCAAACGTAAAGCATTTCCGACATCAAGGCA	2548
682	sCYsThrGLuSerLeuGLuLeuAsp.....	690
2549	ATGTCCTCCCTACCGATCGATTAAGGATTCATTTTGAAMAACGGCGTTT	2598
691ThrLleAlaSerLeuSerVal.....	697
2599	ACCGGACAACCTAGCGCGACGACAGAAACAGCATTTACCTTAAAGAAG	2648
698	ThrLysAsnLleuAlaGLYsGLInGLYGLYLeuHisAlaLys.....	712
2649	CGAATGGACGTCGCGTCAAGGACGAAATTAGCAATTTAAACCTTGACA	2698
713ThreusInLleSerA.....	718
2699	AC.....GCCACATTACACTCATTCCTCCGCTATCGCACGAT	2736
718	sInLeuYsSerGLYLeuSerPheSerAnAsnThrAlaAsnSerSer	734
2737	GCATGACGGCGCCAAAACCGCGAGNCTGTGCAGACAGCCGCCG.....	2782
735	ThrGLYAlaLalArhThrLalArhThrSerGLuSerProThrValSer	751
2783	GTTCCGGCGCGTTCCTATTATCCGTTAACACGCCCAACTCGGTAAATCC	2832
751	rPheLeuPro.....ArgAlaThrAlaGLYSerProAlaP	764
2833	CGTTTAAACACGCTACGCGTAAACGGCAATTTGAACNGTCAAGAACATT	2882
764	roAlaIn.....ThrThr	768
2883	CCCCTTATGTGCGAAGCTTTCGCTACCGAAGCACAATTTGAAGCTGG	2932
769	ProThrTYrAlaGLYAl.....	774
2933	CGGAAGATTCCGAAGNACTTACACCTTGGCGGCTACACAATTCGGGCAAC	2982
775ValGLYGLYAlaLleYrGLY.....	781
2983	GAACCGCTAAGC.....CTGCATCAATTGACGGGTAGT	3014
782	GLuThrValSerPheSerIYsSerGLYLeuCYsGLInPheThrGLuAs	798
3015	GGAAGGAAAGACAACAACCGCTGTCCGAAAACCTTAATTCACCTTCG	3064
798	nserAlaILeAspAsnThrProSerSerProSerLeuAsn.....	811
3065	AAAACGACACGTCATAGCCGGCGGTGGCGTTTACCACACTCATCCGAAA	3114
812ValGLInGLYGLYAla.....LleYrAlaLY	820
3115	GACGGCGAGATTCGCGCTGCATTAATCCGGGCAAGAACAAGAC.....	3157
820	sThrSerLeuSer.....ILeGLuAlaGLuAspProSerThrSerT	834
3158TTTCCGACAAACTCGCAAGGACAGAACCAAAAAACAGCGCGAAA	3202
834	yValPheSerIYsAsnSerVal.....	841
3203	AACACAACGCGCAACCTTTCGACGGCGATTTGGCGCGGCGGCGATCTCC	3252
842SerThrGLYsAlaInThr.....	848

3253 GCGGAAAGACAGAAAGGGTTGGCCGAACCGGGCCGGCGGCGGCGGGG. 3301
849 ThGlyAlnIleAaIyG 855
3302 AAAATGTCGGCATTTATCGAGCGGAGGAAAGAGA 3334
855 lYAlaIeTySerProSerValThrLeuAnaCys..... 866
3335 AAAAAGCGGTGAGCGCGGATAAAGACACCGCGNTTGGCGAAACACCGCGGA 3384
867 GlnThrValPheSerGlyAsnSerAlaIse 876
3385 G.....CGAAACCCGGCGGNTTACCAACCGCTTCCSCCCCGC 3422
876 rMetAlaThrTrAnpProProSerGlyThrSerProLysAspThrIleG 893
3423 CCGCAGCCGCCCGCGGATTTGGCCGAACCGGACGAGCCCAACCGCAACCTC 3472
893 lYcIYAlaIleAlaGlyThrThrIleSerLeuSerLysTyrSerHisPhe 909
3473 AACCCCAACCGGAGCGCGCATGATNACCGCTTATGCAATAGCGGTTTG 3522
910 SerGluAsn..... 912
3523 ACTGAATTTTCCGCAACGCTCAACAGCGTTTTCGCCGTACAGGAGATT 3572
912 912
3573 GAGCCGCGGTGTGGCGAAAGACCGCGGACGAGCGNTTTGGCAAGCNGCA 3622
913 ..ThrAlaAspLeu.....GlyAlaAlaIleGlyThrLeuSer 924
3623 TCCGGAACACCAAACTACCGCTTGGCAAGATTTCGGCGCTTACC... 3667
925 GtYgLYSerSerSerAsnLeuThrGtLYStIleThrLeuSerAsnGlyse 941
3668GCCAACAA.....CCGACCTGGCGCA 3689
941 rPheThrPheGtLYAsnLYAlaAsnLYsArgLYAlaIleTYrAlaP 958
3690 AATCGGATATGACGAA.....ACCTCG 3712
958 roSerValSerIleLysGtLYAsnAlnIleThrPheAsnGlnAsnThrSer 974
3713 GCAAGCGCGCGCTCGGCATCCGTGTTTGGCACACCGGACCGAAACAAAC 3762
975 ThrHisAspGtLYSerAlaIleTYrPhe..... 983
3763 TTTCAGCAGCGGATCGGCACATCGGACGCGGTTCGCCAGCGCGCGTTT 3812
984ThrLysAspAlaThrIleGlnSerLeuGlySerValLeuPheT 998
3813 CCGGCAATACGCGATCGCAGGTTCGACATCGGCATCGCACAGCGGCGCG 3862
998 hrcLYAsnAsnValThrAlaGtLYGlnAlaSerSerThr..... 1010
3863 GTTTTACGAGCGCGCATCTNTACAGCGCGCATCGGAGGAAATCCGCGCC 3912
1011AlaThrGtLYGtLYnThrTrAsnThrTrAsnTYrGtLYAl 1024
3913 CCGGTGCTCATTTACGCGCATTCAGCGACAGATACCGCGCGGTTTCGGCG 3962
1024 aaIa.....IlePheGtLY 1029
3963 AATTCGCGCATGAAACCGGTACA...TGGCGGCAACCGGTAATTTGTCGCAAA 4009
1029 spProGtLYThrGlnThrTrAspThrThrLeuLYsLeuIleAlaSerSer 1045
4010 AACGGATATACCGTATCGGAAAC.....GTCAATATGCGCAC... 4047
1046 GtLYAsnIleThr...PheSerAsnAsnSerGlnAsnThrAlaThrAsnAsn 1061

[illegible]

```

RC STRAIN-IFO 3046;
RX MEDLINE=86223815; PubMed=3011754;
RA Yanagida N., Uozumi T., Beppu T.;
RT "Specific excretion of Serratia marcescens protease through the outer
RL membrane of Escherichia coli.";
RL J. Bacteriol. 166:937-944(1986).
RN [2]
RP PARTIAL SEQUENCE, AND PROCESSING.
RX MEDLINE=92348352; PubMed=1639760;
RA Shikata S., Shimada K., Kataoka H., Hornouchi S., Beppu T.;
RT "Detection of large COOH-terminal domains processed from the
RT precursor of Serratia marcescens serine protease in the outer
RL membrane of Escherichia coli.";
RL J. Biochem. 111:627-632(1992).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M13469; AAA26572.1; -
DR PIR: A29840; A29840.
DR HSSP: Q99405; IMPt.
DR MEROPS: S08.094; -.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; FALSE_NEG.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR HydroLase: Serine protease; Zymogen; signal.
FT SIGNAL 1 27
FT CHAIN 28 645 EXTRACELLULAR SERINE PROTEASE.
FT PROPEP 646 1045
FT ACT_SITE 76 76 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 341 341 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 1045 AA; 112345 MW; 4924EA50E4F179C CRC64;

alignment_scores:
Quality: 177.50 Length: 1190
Ratio: 0.362 Gaps: 56
Percent Similarity: 41.176 Percent Identity: 19.160

alignment_block:
US-09-303-518D-651 x PRTS_SERMA ..
Align seg 1/1 to: PRTS_SERMA from: 1 to: 1045

1135 AGGTAACACACGGT.....GAAACCT 1157
|||||:|||||:.....
91 ArgLeuAspAsnGlySerTyrAsnPheserTyrAspLysGlnAspAsnMe 107
1158 TTCCTTTATGCGATTACGGCAACGCAACATCATCTTATCAACAACATCA 1207
|||||:|||||:.....
107 tserPheGlyAspHisGlyThrHisValAlaGlyIleLeuAlaAlaLysA 124
1208 ACCAAGCGCGCGGC.....GGTTGTATTGTAAGGATTTTACGGTC 1251
|||||:|||||:.....
124 tGAspGlyAlaGlyMetHisGlyValAlaPheAspAlaAspIleLeGly 140
1252 TCGCCTGAAAC..... 1263
|||:
141 ThrIleuAsnAspTyrGlyAsnArgAsnGlyArgGluLeuIleG1 157
1264 .....AACGAACGTGGCAAGCGCGCGCTTCATA 1294

```

```

157 nserAlaAlaArgValIleAsnAsnSerTrp...GlyIleAlaProAspI 173
1295 TCAGTCAGACACAGT.....ACCGTTACTGGAAAGTAAAGGC..... 1332
|||:|||||:.....
173 LeArgArgAspAlaLysGlyAspIleIleTyrLeuProAsnGlyArgPro 189
1333 .....GTGGCAACGACCGCGCTGTGTCAAATCGGCAAGG 1367
|||:|||||:.....
190 AspTyrValAlaPheValLysSerGluValIleAlaGluMetMetArgSe 206
1368 CACGCTGCAGCTTCAAGCCAAAGGGAACCAAGCCTCGATCAGCGTGG 1417
|||:|||||:.....
206 LysSerSerValGlu...TyrGlySerGluGlnProValProThrGlyG 222
1418 CGCAGCGTACATCATTTTGGATCAGCGCAGCAGATGAAGCAAAAAA 1467
|||:|||||:.....
222 LysHisSerAlaMetSerThrLeuArgAlaAlaArgHisGlyLysLeu 238
1468 CAAAGCTTATGTAAGAAATCGCTTGNTCAGCGCAGGAGTACGTTCAACT 1517
|||||:|||||:.....
239 IleValPheSer.....AlaGly..... 244
1518 GAATGCGGAAATACAGTTCAACCCGAC.....AACTCTATTTCG 1558
|||:|||||:.....
245 AsnTyrAsnAsnTyrAsnIleProGluValGlnLysSerLeuProTyrA 261
1559 GCTTCGCGCGCGACGTTTGATTAACGGGCATTCGTTTCGTTCCAC 1608
|||:|||||:.....
261 LysPheProAsp.....ValLeuAsnAsnTyrLeuIle..... 271
1609 CGTATTCAAAT...ACCGATGAGGGCGGATGATTGNCATCAATATATGC 1655
|||:|||||:.....
272 ...ValThrAsnLeuSerAspGluAsnGlnLeuSerValSerSerThrSe 287
1656 C.....ACAACAATCCACCGTTACATTACAGGAATGA.... 1692
|||:|||||:.....
287 rcysGlyGlnThrAlaSerTyrCysValSerAlaProGlySerAspIleT 304
1693 .....AGTATTACCAACGAGGATGATGATGATGATGATGATGATGAT 1728
|||:|||||:.....
304 yserThrValGlyArgLeuGlnSerAsnThrGlyValAlaValAsnArg 320
1729 CTTAATTACAGCAA...GAAATGGCTTACCAACGTTGTTGGCGAGAA 1775
|||:|||||:.....
321 GluAlaTyrAsnLysGlyGlnLeuSerLeuAsnProGlyTyrGlyAsnLys 337
1776 AGATACGACCAAA..... 1788
|||:|||||:.....
337 sserGlyThrSerMetAlaAlaProHisValThrGlyValAlaAlaValL 354
1789 .....ACGAACGGCGGCTCAACCTGTTTAC 1815
|||:|||||:.....
354 euMetGlnArgPheProTyrMetSerAlaAspGlnIleSerAlaValIle 370
1816 CACCCCGCGCAGACAGACCGCACCCNGCTGTTTCGGCGGACAATTT 1865
|||:|||||:.....
371 LysThrThrAlaThrAspLeuGly.....ValAlaGlyIleAspAsnLe 385
1866 AAACGCG.....AACATCAGCAACA...AACGCG...AACG 1897
|||:|||||:.....
385 urPheGlyTyrGlyArgValAsnLeuArgAspAlaIleAsnGlyProLysM 402
1898 TGTTTTCAGCGCGCAGACGACGACGCGCTTACATATTTAGGAACG 1947
|||:|||||:.....
402 etPheIleThrLysGlnAspIleProGlnGlnLutyr..... 413
1948 GGGTGTCTCAAAATGGAAGGTATCCCAAGGAGAAATGCTGTGGACAA 1997
|||:|||||:.....
413 ..... 413
1998 CGACTGATCNCAGCAGCTTAAAGCGAAATTTCCATTATTCAGGGCG 2047
|||:|||||:.....

```

```
414 ....TyrValProGlySerTyrSerGluLysGlnPhe..... 424
2048 GGCAGCGCGTGAATTCGCCCAATGTTGCCAAGTGGAGGCGATTGNCAT 2097
425 .....ValValAsnIleProGly..... 430
2098 TTGAGCAATCAGCGCCCAAGCAAGTTTGGTGTGGACCGCATCAAGGCA 2147
431 LeuGlyAsnIleValGlnProGlyThrProValGluArgGlyThrSe 447
2148 TACAACTGTACAGCTTGGAGTGGACGACGNGTGTGCAAAATTGTGCGA 2197
447 TserGlyCysSerPheAspSerTyrPser..... 456
2198 AAACATTACCGACGATGAAGTGAATGATTCATGCTACATAAGCAGCAT 2247
457 .....AsnAspIleSerGlyHisGlyGlyLeuThrLysThrGly... 469
2248 AGCGGCATGTAGCNCNTNCCNATNAGCNTNTNAAANCTCNCNGGCG 2297
470 .....AlaGlyTh 472
2298 TGCNCACTNANAGCAATCTTAGTGAACATGCGATACAGCTTATACAG 2347
472 ThrValLeuLeuGlyAsnAsnThrTyrArgGlyAspThr..... 485
2348 TCAGCCACACGCGCACCAAAAGCGCAACCTTAGCCTGTGGGCAATGCC 2397
485 ..... 485
2398 CAGCAACATTATTAACAGCCATTAAGCGCAACNCATCGGNTTTCGGG 2447
485 ..... 485
2448 CAATGCTCATTTATTAATGAACACAGCGCGCAAAAGGCAAGTGTGA 2497
486 .....TyrValLysGlnIleValLeuA 493
2498 CGCTTCCGACACGCGTAAGGCAAGCAAGTGAAGCATTCGCACTAACGCG 2547
493 IalIeAspGlySerValAlaSerAsnVal..... 502
2548 AATGCTCCCTAGCGGATGAAGGCAAGTATTCATTTTGAACACAGCGCTT 2597
503 .....TyrIleGluAsnSer..... 507
2598 TACCGACAACTCAGCGCGACAGCAAGNACAGCAATTACATTAAGAACA 2647
508 ....GlyThrLeuSerGlyGluGlyThrValGlyAlaPheArgAlaAla 523
2648 GCGAATGACGCTGCGCGTACAGGCAAGCAATTAAGCAATTTAAAC...CTT 2694
523 TserGlySerValAlaIleProGlyIleGlyThrLeuIleValLeu 539
2695 GACAACGCGCCATTAACATTAATTCGCGCATGCGCGCAAGTGTGTGCA 2747
540 HisAspAlaIlePheAspArgGlySerGlnTyrAsnValGluValAlaAs 556
2745 CGCGCAACCGCGCAGNGTGTGACAGACGCGCGCGCGCTTCCGCGCGCTT 2794
556 P..... 556
2795 CCCTATTATTCGTTACACCGCAACTTCGCTAGTAATCCGTTTCACAGC 2844
556 ..... 556
2845 CTGACGGTAACGGCAATTAACNGTCAGAGACATTCGCTTTATGTC 2894
557 .....AsnGly..... 558
2895 GGAACCTTTCGCTACGCAAGGCAAAATTAAGTGAAGTGGCGAAAGTTCG 2944
559 .....ArgSerAspLysIleAlaAlaIleArg..... 567
2945 AAGNACTTACACCTTGCGGTCAACAATACCGCAACGAAACCCGTAAAC 2994
568 .....AlaPheLeuAsnGlnIleGlySerValAsnValSer 578
2995 CTGATCAATTAAGCGTAGTGAAGGGAAGACAAACAAACCGGTGTCCGA 3044
579 LeuGluArg.....SerG 583
3045 AAACCTTAATTTACCTGCAAAACGAACAGTCGATCCGCGCGGTGCG 3094
583 AsnLeu.....LeuSerGlnAsnGlnAlaGlnSerLeuLeuGlyAsnI 598
3095 GTTACCAACATCATCCGCAA.....GACGCGAGTTCGCGCTG 3132
598 YstIrrThrIleLeuThrThrAspGlyValThrGlyArgPheGluAsn 614
3133 CATTAATCCG.....GTCAAGAACACAGAGCTTTCGCA 3164
615 AlaAsnProSerTyrProPheValLysValAlaLeuAspTyrArgGlyAs 631
3165 CAACCTGCGCAAGGCAAGCAACCAAAACAGCGCGGAAAAAGCAACGCGC 3214
631 AspValGlyLeuGlyLleThrArgThrAspAlaSerPheAspSerLeuA 648
3215 AAAGCCTTGACGCGCTGATTGCGCGCGCGCGCATGCCCGCGCAAAAGCA 3264
648 IAserThrGluAsnGlnLysAlaValAlaArg.....AlaValGluThr 662
3265 GAAAGCGTTGCGCAACCGCGCGCGCGCGCAAGCGGGAATTCGCGCAT 3314
663 LeuAsnAlaThrGluProValThrGluThrAlaLysArgSerValAlaI 679
3315 TATGACGCGGAGAGAGAAAAACGGGTGCAGCGCGGATTAAGACAGCG 3364
679 e...ProAlaIleGluIleValAsnLeuLeuGlnSerAspGlyGlyAla 695
3365 CATTGCGCAACAGCGCGCAAGCGGAAAC.....CGCGCGGTACG 3405
695 IeGlnAlaValAsnGlnGluIleAsnIleValAlaGlnIleProIleTyr 711
3406 ACCGCTTCCCGCGCGCGCGCGCGCGCGCGGATTTGCGCGCAACGCGA 3455
712 GluSerPheLeuGlyPheThrSerAlaArgL.....LeuGlnI 725
3456 GCCCAACCGCAACGCTCAACCCCAACCGCGCGCGCATTAACCGCTT 3505
725 AlaThrArgGlnLeuSerGlyGlnIleHisAlaAspMetAlaSerAlaG 742
3506 ATGCAATACGCGGTTGAGTGAATTTCCGCCACGCTCAACAGCGCTTTC 3555
742 IlnIleAsn.....GluSerAlaGlyT 748
3556 GCCGTACAGAGCAATTAAGACGCGGTG...TTTCCGAAGACCGCGCGC... 3600
749 LeuArgAspThrAlaThrGluArgLeuArgGlnAlaGlyArgArgTh 765
3601 .....AACGCGTTTGGACAACGCGCATCC 3625
765 AlaThrAspIleLysAlaAspAsnGlyAlaThrAlaLysLeuLeuG 782
3626 GGNACACCAACACATACGTTGCAAGAT...TTCCGCGCGCTACCGCA 3672
782 LysThrProGlnHisAlaSerGlyAsnAspAsnAlaThrGlnIleArgGlnThr 798
3673 CAACGCGACCTGCGCCAAATCGGTATGCAAAAAACCTC...GCGACGCG 3719
799 SerThrTyrGlyValLeuLeuGlyLeuAspSerGlnLeuPheGlyAspG 815
3720 GCGCGTGGGCAATCCGTTTTCGACACACGCGCAACCGCAACCACTTCGACG 3769
815 YArgLeuGlyMetMetThrGlyTyrThrArgThr.....SerLeuAspG 830
```

```
3770 ACGGCATCGGCACTCGGCAAGGCTTGCCACGCGGCCGCTTTGGGGCAA 3819
      ||| .....|||::: ::::: |||
830 1yGlyTyrGlnSerAspAlaHis...SerAspSntYrHisLeuGlyLeu 845
      |||  |||
3820 TACGGCATCGGCAAGTTCGACATCGGCATCGACGCGGCGGTTTAC 3869
      |||  |||
846 TyrGlyAspLysArgPhe..... 851
3870 CAGCGGCANTCTNTCAGACGCGCATCGAGCGCAAAATC..... 3906
      |||  |||
852 ....GlyAlaLeuAlaLeuAlaGlyGlyTyrTyrThrPheHisArgI 867
      |||  |||
3907 ....CGCGCGCGGCTGCTGATTCAGGCATTCAG..... 3936
      |||  |||
867 leAspThrSerArgSerValAsnYrGlyAlaGlnSerAspArgGlyLys 883
3937 GCACGATACCGCGCGCGGCTTCGGGGATTCGGCATCGAACCGTACATCGG 3986
      |||  |||
884 AlaLysTyrAsnAlaArgThrGlnGlyLeuPheIleGlu..... 896
3987 CGCAACCGCGCTATTGCTCCAAAAGCGGATACCGCTACGAAACGTCA 4036
      |||  |||
897 .....SerGlyTyrAspTrpTrpThrSerAspAlaValA 907
4037 ATATCGCCACCCCGGCTTGCTTCAACCGNTACCGNGCGGCATTAAG 4086
      |||
907 sn..... 907
4087 GCAGATTATTTCATCAACCGCGCAACACATNTCCATCANCCTTATTT 4136
      |||  |||
908 .....LeuGluProPheAl 912
4137 NAGCGTGTCCCTATACCGATGCCGCTTCGGGAAAGTCCGAACACGCGTCA 4186
      |||  |||
912 aAsnLeuAlaTyrThrHisTyrArgAsnGluGluIleAsnGluGlnGly 929
4187 ATACCGCGNGTATTG.....GCTCAG 4206
      |||  |||
929 1yAlaAlaAlaLeuAlaGlyLysPlyGlnSerGlnSerAlaThrAlaSer 945
4207 GATTTCGGCAAAACCGCGCATGCGGATGGGCGTAAACCGCGAAATCAA 4256
      |||  |||
946 ThrLeuGlyLeuAlaArgAlaAspThrGluTrpGlnThrAspSerVal..... 960
4257 AGGTTTACGGCTGTCCNTCCACGCTGCCCGCGCAAGNCCGCACTGG 4306
      |||  |||
961 .....AlaIleAlaLeuAlaGly...GluLeuG 969
4307 AAGCGCAACACAGCGCGGCGC 4326
      |||  |||
969 1yTrpGlnHisGlnTyrGly 975
```